

Promises and threats of the knowledge-based economy

The need to place scientific knowledge at the heart of economic and social policy has underlined some of the limitations of focusing on trade liberalization. Developing countries should be a prime beneficiary.

Suddenly everyone is talking about it. For years, the idea that knowledge in general — and scientific knowledge in particular — plays a role in the global economy comparable to more tangible forms of resources and capital has been confined largely to the writing of left-leaning sociologists and policy analysts. As such, it has often been dismissed in political circles as little more than a woolly minded bid for extra funding from the scientific community and tax breaks on R&D spending from industry. Over the past year, however, the concept that we live in an age when both the social and economic health of any society depends critically on its ability to harness scientific (and other) knowledge has gained wide political currency.

For some countries, of course, there is little new in the idea of using technical knowledge as a source of wealth and power. Those facing high labour costs (such as the United States) or a relative lack of natural resources (Sweden, Switzerland or Japan) have long realized that superior technical skill and innovation are their key to maintaining a competitive edge in the global market-place, and have sought to structure their industrial and social institutions around such insight. US economists such as Edwin Mansfield and Richard Nelson have identified the critical contribution made by technical innovation to economic growth; a recent analysis by 'knowledge management' expert Karl Sveiby suggests that half of the fastest-growing US companies are 'knowledge companies', selling the knowledge and skills of their employees rather than manufacturing products or providing services.

Elsewhere, for example in some of the major European economies, the message has taken more time to sink in. In the past, the prime role of scientific knowledge in the economy was portrayed as largely instrumental, leading to new products or reduced production costs. Only recently has the strategic value of intangible assets — such as scientific skills or intellectual property — come to be fully acknowledged. This was the key theme in the British white paper on competitiveness, published last month (see *Nature* 396, 714; 1998). It has also come fully to the fore in the European Union's fifth four-year Framework programme, due to be launched this year.

Conversely, part of the blame for the recent collapse of confidence in various Eastern economies is linked to their failure to establish a secure knowledge base for their economic 'miracles'. Proper information also oils the wheels of the market; when it is lacking, growth and investment can grind to a halt. But as acknowledgement of the significance of living in knowledge-based economies has spread, so too has experience of its drawbacks and limitations. Increased awareness of the potential economic value of scientific knowledge has led to growing reluctance to make such knowledge freely available throughout the research community. And the use of data banks to monitor the behaviour of individuals or market new products, both knowledge-intensive activities, has raised unprecedented questions about human privacy — including most recently those concerning an individual's genetic identity.

Last, but by no means least, as a recent report from the World Bank has pointed out, one of the most dangerous aspects of the global

knowledge-based economy lies in the tensions created by the growing 'knowledge gap' between the knowledge-rich countries of the North and the (generally) knowledge-poor countries of the South (see *Nature* 395, 529; 1998). Wealth creates the ability to create the knowledge that can be used to create further wealth. But, without adequate means to distribute the benefits accruing from such knowledge, social disparities, and the jealousies they invoke, will only increase.

The key lies in combining commitment to two concepts that are far easier to define than to achieve: scientific excellence and social equity. The first requires continual vigilance and, occasionally, hard choices; without these, there is a tendency (as with currencies) for the mediocre to drive out the good. It is no longer sufficient to justify a research programme purely in terms of a generalized commitment to human welfare, just as a country (or a company) cannot guarantee itself economic success by leaping on the latest technological bandwagon. In each case, whether one is talking of scientific results or technological products, the combined effects of modern communications and trade liberalization mean that the closer these meet specific needs and global standards of quality (and cost-effectiveness), the greater benefit they are likely to generate.

But this does not mean that an untrammelled free market in the spread and application of scientific ideas is the most effective basis on which the knowledge economies of the future will be built. Governments may no longer have a role in picking technological winners, but they still have an essential task in both providing a sound underpinning for efforts to generate and exploit knowledge opportunities and compensating for market weaknesses and failures. Part of this responsibility lies in ensuring that moves to address needs not directly catered for by the market economy are nevertheless placed as firmly on a healthy scientific footing as are wealth-generating activities.

There have recently been encouraging signs that this message, too, is getting through to policy-makers, particularly in the context of aid to poorer nations. As shown in a series of articles in this issue by *Nature* correspondents around the world, the limitations of aid strategies that focus on providing external technical assistance or altering purely economic mechanisms are becoming widely recognized (see page 6). The key to successful aid is to find ways of enabling poorer countries to make their own way in the global knowledge economy — and of removing obstacles while at the same time harnessing knowledge-generating capacities to the countries' social needs. This has been made both simpler and more urgent by the arrival of the Internet.

Devising strategies to achieve both objectives is likely to dominate the agenda of the World Conference on Science, to be held in Budapest at the end of the June, jointly organized by the United Nations Educational, Scientific and Cultural Organization and the International Council for Science. To accompany the preparatory discussions, *Nature* is dedicating part of its website (www.nature.com) over the next six months to detailed coverage of and comment on the challenge ahead. We encourage readers to participate in the debate. A constructive outcome would be a fitting way to greet the new millennium. □



US universities create bridges between physics and biology

[WASHINGTON] A number of leading US research universities are planning new institutes to bring physical and biomedical scientists together. This reflects a growing feeling that these fields should be linked more closely in both research and teaching.

One of the largest initiatives comes from Stanford University, where the Nobel-prizewinning physicist Steven Chu and biochemist James Spudich are spearheading a proposal for a research centre housing 50 faculty members, spanning disciplines from applied physics to clinical medicine.

Multidisciplinary research centres are also planned at the University of Chicago, which is setting up an 'interdivisional institute' straddling the biological and physical sciences, and the University of California at Berkeley, which is planning a building for its bioengineering department and some faculty members from molecular biology and several physical science departments.

In addition, Princeton University is due to announce plans this week for an interdisciplinary genomics institute in a new \$40 million building connected to its molecular biology department.

About 10 of the 50 posts in the Stanford institute would be new positions. The proposal, which the university administration supports, calls for a new building of 200,000 square feet, which would make it one of Stanford's largest research centres.

University staff declined to estimate the project's cost. But the building alone would cost "tens of millions of dollars", according to Charles Kruger, dean of research.

Although some hurdles — including raising outside funds — remain to be cleared



before the centre can receive formal approval from Stanford's board of trustees, its supporters expect the plan to go ahead, and hope to have the new building by 2002.

The project is currently called 'Bio-X', reflecting the desire to mix biologists with researchers from other disciplines, but without an explicit agenda of techniques to be used or problems to be solved.

"We just want to mix smart people together in an interdisciplinary environment, and let nature take its course," says Spudich. The new building will be between the medical school and the science departments, and Spudich hopes it will draw people in from the surrounding buildings.

The promise of the centre has already attracted one prominent new faculty member to Stanford. Princeton biophysicist Steven Block, a pioneer in the use of 'optical tweezers' to study molecular motors, has accepted a joint appointment in Stanford's

applied physics and biology departments, starting in September.

Bio-X is motivated partly by the growing perception that a deeper understanding of complex biological systems will need a more quantitative type of biology that is closely integrated with the physical sciences.

Much of molecular biology already relies on experimental techniques invented by physicists, such as NMR and X-ray diffraction. But as biology becomes more data-rich, it increasingly requires the analytical and computational methods characteristic of the physical sciences.

At the same time, physicists are finding new problems in biology. For example, Chu, who won the Nobel prize for work on laser cooling of trapped atoms, now also works on the behaviour of single protein molecules.

The desire to bring together scientists from different disciplines stems from an awareness that traditional departments can be physical barriers to cross-fertilization. This is part of the thinking behind Princeton's new institute, planned to be built within two years, with about 12 faculty drawn from physics, chemistry, mathematics and engineering, as well as from biology.

Developmental geneticist Shirley Tilghman, the institute's director, says it will focus on what she calls "integrative biology" — the integration of many kinds of information to understand complex biological processes, such as interactions between cells or genes.

"Biology has thrived in the past 50 years by taking things apart and identifying their components," says Tilghman. "But there is a consensus in the field that the time is coming very soon to reverse the process" — to study how these components interact to form complex systems.

Chicago approved its interdivisional institute in June 1997, and two co-directors were appointed last year. A building of more than 200,000 square feet is planned, costing \$110 million and due to be built by 2002. It will house not just the 24 faculty members of the new institute (half of whom will be new appointments), but also several existing departments and Howard Hughes investigators.

The new centres should each have an impact on graduate and undergraduate teaching, as well as research. "There is a growing feeling in the biological community that we need to be thinking hard about how to train the next generation of biologists," says Tilghman. She suggests that this training should include more mathematics, physics and chemistry.

Laura Garwin

France questions Israeli research deal

[JERUSALEM] Israeli officials have condemned a French-led effort to make Israel's participation in the European Union's fifth Framework research programme conditional on it implementing the 'Wye Plantation' agreement reached with the Palestinian Authority in October.

Pierre Lebovics, spokesperson for the French embassy in Tel Aviv, says that, although Israeli participation is desirable in principle, "it is clear that such participation has a political dimension, in the context of furthering the peace process and the implementation of the Wye accords".

But Orna Berry, Israel's chief scientist, describes France's stance as "a flagrant violation of professional integrity". She says

she will refuse to co-ordinate Israel's participation in the Framework programme unless an agreement with Israel is signed by the European Union within two weeks.

Berry says that, if the matter is not resolved by mid-January, when the first calls for projects are issued, there will be little sense in Israel participating.

She argues that Israel's strong research community, and her own experience in working with the Palestinians and Jordan, can contribute to European research and development. "The right thing to do is to concentrate on what we have in common and not on what separates us." The issue will be taken up by Europe's Council of Ministers early this month.

Haim Watzman

Eastern Europe still strapped for cash...

[MUNICH] None of the central-east European countries that are candidates for membership of the European Union (EU) plans to significantly increase its research spending next year, despite announcements that they would do so in anticipation of joining the EU.

The European Commission is concerned that, unless the governments of Poland, Hungary, the Czech Republic and Slovenia increase their spending soon, they will be badly placed to exploit their new status as full associates of the commission's fifth Framework programme of research (FP5), which will be launched this year.

This status, which allows them to participate in all FP5 programmes in exchange for a financial contribution equivalent to that of EU member states, is intended to help introduce the countries into the culture of the EU.

Along with Estonia and Cyprus, the four countries are currently negotiating EU membership after 2003. But each spends much less on research and development than the EU average of 1.8 per cent of gross domestic product (GDP), equivalent to US\$350 per capita.

A major criterion for membership is that their economies should be roughly aligned with EU norms, and the commission has been encouraging each candidate to raise its research spending as part of this effort.

"The average expenditure in candidate countries is much lower than that in EU members, although there is some overlap in the ranges of expenditure," says Rainer Gerold, head of the commission's section on research and eastern European countries.

Slovenia, for example, already spends more than some of the poorer member states, which are also a worry to the commission. In 1995, the Slovenian government and industry's combined contributions added up to \$150 per capita (1.7 per cent of GDP), compared with Greece's \$60 per capita (0.7 per cent of GDP).

But, in contrast, Poland and Hungary spent only \$21 (0.7 per cent of GDP) and \$28 (0.7 per cent of GDP) per capita, respectively, in the same year, and the Czech Republic spent \$50 per capita (1.0 per cent of GDP).

No fixed targets for research expenditure have been defined for the candidates, says Gerold, but discussions have led to governments setting their own goals. According to government statements, Poland, Hungary and the Czech Republic are aiming for public expenditure alone on research of 0.7 per cent of GDP before 2003, and Slovenia is aiming for 1.2 per cent of GDP. Each expects its industry to match public spending.

But scientists in these four countries are sceptical about the seriousness of these intentions, as promises to begin significant spending increases in 1999 have failed to

materialize in terms of firm commitments.

Indeed, in Poland a significant cut in spending was avoided only after a dozen or so top scientists formed a group called Save Polish Science and orchestrated a vigorous press campaign against the government's proposed cuts. As a result, public spending on research and development will remain at around 0.45 per cent of GDP next year, where it has stagnated for most of the 1990s.

Hungarian scientists also had to fight to maintain present spending levels. According to Laszlo Keviczky, president of the Hungarian Academy of Sciences, scientists "were lucky to emerge with a steady-state budget which takes into account inflation".

Rudolf Zahradník, president of the Czech Academy of Sciences, foresees no significant rise in the Czech science budget this year. He is angry with the government for not giving a clear indication of how the target for the period before EU accession will be reached.

Slovenian scientists fared better. "With a seven per cent rise in R&D budget in real terms, we cannot complain so much," says Venčeslav Kaučič, head of inorganic chemistry at the National Institute of Chemistry in Ljubljana. But he points out that the past three years have seen a significant decline in government spending on research from 0.8 per cent of GDP to around 0.7 per cent. Spending power has remained roughly constant because of the growth in GDP.

Although promises for increased funding in 1999 have been broken, the next set of promises is already being made. Poland's vice-president Pavel Mordlyk, who chairs the government's science and technology advisory committee, announced last month that "sub-

stantial increases would start in 2000 or 2001".

Candidate countries began negotiating the details of their full associate membership of FP5 last October. In order to help them readjust their overall finances, they will be offered decreasing annual discounts on their contributions to the FP5 budget, with the full contributions being payable in 2002.

European Commission officials admit they are worried that, if countries do not increase spending on domestic research and development soon, they will find difficulties meeting the later FP5 payments.

Lack of research investment is also threatening the future of research in these countries by keeping salaries in the public sector low, and by limiting the amount of internationally competitive work that is possible.

Andrej Wroblewski, a professor of physics at Warsaw University and a member of Save Polish Science, says that the situation for scientists in Poland is "desperate". "Science is collapsing here," he says, pointing to a dearth of PhD students — the number has fallen to its level 30 years ago — and a slow erosion of Poland's position in the world's publication ranking list.

The prospect of higher financial rewards is turning the smartest graduates away from research and towards sectors such as commerce and banking. "This internal brain-drain is robbing our countries of our next generation of scientists," says Wroblewski.

An external brain-drain is also causing problems. A recent evaluation of Hungarian biology by the European Molecular Biology Organization (see below) concluded that low salaries are still driving the best scientists to emigrate.

Alison Abbott

...but much Hungarian biology is world class

[MUNICH] Biological research at institutes of the Hungarian Academy of Sciences is generally of a "high standard", and one-third of the research groups are of international standard, according to an international evaluation committee.

But the committee, which was chaired by Kai Simons, head of cell biology at the European Molecular Biology Laboratory, warns the Hungarian government that standards will collapse if investment in research continues at a low level.

The evaluation was requested by the academy last year and organized by

the European Molecular Biology Organization. The committee of 12 scientists from around Europe judged 16 out of the 46 groups it assessed — predominantly in plant biology and developmental genetics — to be "world class".

The committee said this was a remarkable achievement given the low level of funding. But it warned that "it will not be possible to maintain this level of excellence in the future if funding is not increased".

"There are already some visible signs of deterioration which do not bode well for the future," it said. The panel

found that the best scientists are still going abroad, leaving behind a lack of potential young group leaders, while the standard of equipment is generally low.

The two most urgent problems, it said, are the very low wages and the small size of the average research grant. With salaries ranging from US\$240 to \$660 per month, most researchers have to take second jobs.

The committee recommends that the government establishes a programme of generous grants to allow young researchers to work to international standards. **A.A.**

Japanese budget boosts science funding

[TOKYO] Japan's largest ever budget, aimed at reviving the nation's ailing economy through increased public spending, includes a generous 8.1 per cent boost in expenditure on science, with a strong emphasis on basic research.

The budget for the 1999 fiscal year, which begins on 1 April, takes overall science spending across all ministries and agencies to ¥963 billion (US\$8.5 billion), with large increases for life sciences in particular. Also featured are support for creating new businesses through 'commercially applicable research', and the promotion of collaborative research between universities and industry.

The total government budget of ¥81.9 trillion — an increase of 5.4 per cent from the previous year and the highest growth rate in 20 years — was approved by the Cabinet on 25 December. The increase in science expenditure is in line with the five-year plan for science and technology, under which the government pledged to double the nation's research spending between 1996 and 2001.

The budget for the Science and Technology Agency (STA) has grown by 4.1 per cent from last year, as a result of increased support for brain and genome research, especially with last year's opening of the Genome Sciences Centre, backed by the Institute of Physical and Chemical Research. Spending on nuclear and space research was kept down so more funding can be allocated to basic research in other areas, according to STA.

Spending on space development will increase by 2.6 per cent by the last-minute addition of funding for a reconnaissance satellite to improve Japan's ability to gather security information, following last year's firing of a North Korean missile over Japanese territory (see *Nature* 396, 401; 1998).

The satellite will be allocated ¥11.3 billion in 1999, with ¥6.8 billion going to STA

(included in the space development budget), ¥2.2 billion to the Ministry of International Trade and Industry (MITI), ¥1.4 billion to the Cabinet secretariat, and ¥0.9 billion to the post and telecommunications ministry.

STA will also gain an extra ¥2.7 billion to produce SELENE, a lunar probe jointly developed by MITI and the Ministry of Post and Telecommunications. Although the overall budget for the Ministry of Education, Science, Sports and Culture (Monbusho) has increased by only 1.4 per cent, its spend-

ing for scientific research has grown by a comparatively generous 6.7 per cent.

The Long Baseline Neutrino Oscillation Experiment, carried out jointly by the Institute of Cosmic Ray Research at Tokyo University and Monbusho's High Energy Accelerator Research Organization (KEK), will gain ¥1 billion, and high-energy accelerator research at KEK will receive ¥8.5 billion. This is an increase of 372 per cent from last year, to cover the initial operating costs of Belle, a precision particle detector for studying B-mesons (*Nature* 396, 612; 1998).

The postdoctoral fellowship scheme under Monbusho, which aims to increase the number of Japan's postdoctoral researchers to 10,000 by the millennium, is expected to reach its target this year. Monbusho and STA will support an extra 1,304 postdocs in 1999, making 10,076 by the end of the fiscal year.

MITI's spending on research has gone up by 8.9 per cent from last year. Support for new venture businesses will receive a spending boost, especially at universities, following the passing of a law promoting collaboration between universities and industry.

The budget has yet to be approved by the Diet (Japan's parliament), but changes to science-related budgets are usually minimal.

This year's dramatic spending increase is in sharp contrast to last year's austere budget. This was drawn up under the Fiscal Structural Reform Act, which sought to cap spending commitments in order to control government deficits and promote budgetary restraint; the law has since been frozen.

While the 1999 budget is aimed at producing the government's projected economic growth — 0.5 per cent in gross domestic product by the next fiscal year — analysts are critical that it will be partly funded by ¥31 trillion in government bonds, putting the nation deeper in debt.

Asako Saegusa

Highlights of Japan's Science and Technology Budget 1999 (in billion yen: US\$1 = ¥113)

Science and Technology Agency (STA)		
Total budget	770.7	+4.1 %
General R&D budget	612.2	+4.6 %
Genome research	11.9	+70.9 %
Brain research	18.3	+30.8 %
Deep-sea drilling	3.4	+453 %
Atomic energy	187.2	-1.9 %
Space development	187.2	+2.6 %
STA fellowships	3.8	+5.5 %

Ministry of Trade and Industry (MITI)

Total R&D budget	508.1	+3.1 %
Agency for Industrial Science and Technology	135.8	+6.7 %
New Industry R&D	44.3	+44.6 %
Joint Research with universities	3.6	+58.9 %
Energy	49.1	+7.0 %

Ministry of Education, Science, Sports and Culture (Monbusho)

Grants in aid for scientific research	187.9	+6.7 %
Japan Society for Promotion of Science	47.3	+12.4 %
(postdoctoral fellowships)	(17.6)	+9.7 %
Joint research with industry	117.8	+11.1 %
Centre of excellence programme	73.8	-5.1 %
High-energy accelerator	8.5	+372 %
Neutrino oscillation	1.0	New

Software problem delays US spacecraft's meeting with asteroid

[WASHINGTON] The Near-Earth Asteroid Rendezvous (NEAR) spacecraft, which was supposed to have begun orbiting the asteroid Eros next week, will meet up with it in February 2000 instead. The delay, which project managers say will not reduce the mission's scientific return, was caused by a software problem when the spacecraft's main engine was fired on 20 December.

NEAR — operated for the US space agency NASA by the Johns Hopkins University Applied Physics Laboratory (APL) — was designed to orbit the asteroid, photographing its surface, mapping minerals, and returning gravity data.

The manoeuvre on 20 December was intended to position the spacecraft to begin

orbiting Eros. But when the engine fired, the acceleration triggered software designed to protect the spacecraft from uncontrolled motion. NEAR automatically aborted its engine burn and waited for a signal from the ground, ruining chances of a rendezvous with Eros this year.

Engineers switched to a contingency plan for an engine firing on 3 January and a rendezvous 13 months later. Another option, for a meeting with Eros in July, was scrapped because the command sequences had to be revised too quickly.

Tom Coughlin, the NEAR project manager at APL, says the software that triggers the abort modes will be reworked to avoid a similar mishap in the future. The

spacecraft wasted about 30 per cent of its fuel during the aborted manoeuvre, but Coughlin says it still has enough to execute the full year-long study of Eros.

Managers want to keep open the option to land the spacecraft briefly on Eros after the main mission ends. Such an extension to the mission, which would have to be approved and funded by NASA, is still possible with the amount of fuel remaining.

The delay is likely to cost NASA millions of dollars, but has at least one benefit. Because the spacecraft was able to return images and take data on 23 December, scientists will have a better understanding of the asteroid's shape and mass when close manoeuvres begin next year. **Tony Reichhardt**

Taking knowledge to the poorest

Three months ago, the World Bank warned of the dangers of a growing 'knowledge gap' between the rich and poor nations (see *Nature* 395, 529; 1998). In the following pages, *Nature's* correspondents describe the increasing attention being paid to the search for ways in which this gap might be bridged.

The first article outlines the origins of the World Bank's concerns, and some of the debates this has provoked inside the bank about the nature of development aid. This is followed by a description of the growing interest in the United States and Europe in placing such aid on a more scientific footing.

Our Indian correspondent describes some cautionary tales of the pitfalls associated with imposing science-based solutions on problems where the social environment has not been adequately taken into account. Finally, we look at the promise offered by new communications technologies to the Third World, in particular the growing use of the Internet.

The picture that emerges is one of promise and opportunity, but also of significant hurdles, both financial and institutional. Both sets of issues will be discussed at the World Conference on Science, organized by the United Nations Educational, Scientific and Cultural Organization (Unesco) and the International Council of Science, and to be held in Budapest at the end of June.

In a bid both to contribute to and to stimulate discussion of the issues, a section of the *Nature* website will follow and comment on the preparations for the conference. The site opens today (7 January) with a contribution from Federico Mayor, the director-general of Unesco, on why science needs to renew its 'social contract' with society, and includes fuller versions of the five articles published here. It can be accessed on www.nature.com. **David Dickson**

World Bank invests in global science base

[LONDON] After 50 years of paying for roads, power and schools, and helping poor countries to liberalize their economies, the World Bank — the financial arm of the United Nations system — has started shifting some of the focus of its activities to supporting 'knowledge development', including science.

Two separate internal World Bank task groups are investigating a potential role for the bank in supporting science in developing countries. Each group will report back this year with proposals on how the bank can best support basic research, something it has never before considered, how to make its expertise more available to developing countries, and whether it needs a science department to oversee its new initiatives.

The bank, which is owned by 180 governments, provides long-term loans at commercial interest rates, mainly to developing countries. One quarter of its lending is interest-free and goes to the poorest. In the 1980s, with its focus on infrastructure development and trade liberalization, it closed its science department and abolished the science adviser's post.

Direct support for research in developing countries is now seen as more of a priority. This is because the bank believes research will help to find solutions to its priority issues, such as providing the poor with access to food, clean water and a disease-free environment.

But it also comes from a belief that developing countries need to build up knowledge-based industries to remain economically competitive. In an attempt to help the poorest countries, particularly those in Africa, to catch up with those better off, the bank is helping to fund information technology infrastructure under a programme called *infoDev*.

As a sign of this new thinking, the bank devoted the latest edition of its annual World Development Report to bridging the 'knowl-

edge gap' between rich and poor countries. Last month it agreed to partly fund in Chile the first in a chain of centres of excellence in scientific research — known as Millennium Institutes — in developing countries (see *Nature* 395, 529 & 396, 711; 1998).

Both events represent the culmination of a three-year study by the bank into how it can fund science in developing countries in partnership with governments and philanthropic foundations. Ian Johnson, the bank's vice-president for environment, acknowledges that the bank previously considered research to be a luxury for developing countries. But he says that attitudes have changed.

"Development in the next 20 to 30 years is going to be more science-based," says Johnson. For example, he says, biotechnology and climate change will have a major impact on world agriculture. "We need to recognize, understand and be prepared for this."

Despite its previous reluctance to provide significant funds for research, the bank is no stranger to science. For example, it hosts the secretariat of the Consultative Group on International Agricultural Research, a network of 16 agricultural research centres mainly in developing countries. It is also one of three partners in the Global Environment Facility, the US\$2 billion United Nations funding agency for environmental projects.

When the bank closed its dedicated science department it decided to place greater priority on encouraging countries to liberalize their economies. This stemmed from the belief that a favourable economic climate was needed for projects to function and loans to be repaid.

Charles Weiss, programme director of science and international affairs at Georgetown University in Washington DC, was the bank's science adviser in the early 1980s. He believes this latest attempt to get the bank thinking about science has more chance of succeeding than his own efforts. This, he says, is partly because they have the support of senior executives, particularly James Wolfensohn, the bank's president, and partly because the bank is now keen to promote knowledge-based development.

This strategy is based on the bank's need for a new role now that private capital has replaced development aid as the main source of external finance for developing countries. Whereas the bank's lending has remained at around \$20 billion for the past five years, private-sector foreign investment in devel-

Gift of knowledge: information technology has an essential role to play in world development.



MARCUS ROSE/PANOS

oping countries has increased sevenfold, to more than \$150 billion per year.

Wolfensohn is known to be keen for the World Bank's profile as a 'knowledge bank' to be raised. "When the bank was a major financial influence, its information role was downgraded," says Weiss. "Now it is the reverse."

Unusually for a lending institution, the World Bank possesses world-class expertise on the projects and the regions where it lends money. Of its 8,000 staff, 3,000 have a PhD-level qualification, and many of these are top-ranked researchers headhunted from universities. The quantity and quality of the bank's research is consistently high.

But this more analytical aspect of the

bank's work has always been overshadowed by its lending arm — known as operations — which has generally considered research to be a function of lending, rather than an activity in its own right. In 1987, half of the research staff were sent to work in operations.

This tension between the research and lending wings remains, and is one of several challenges that will need to be overcome if the new strategy is to bear fruit. In particular, the need for a new department for science is being questioned by some who do not want to see science confined to a ghetto and think it should be part of the lending portfolio of all of the bank's departments.

Some operations staff have yet to be con-

vinced of the merits of raising the bank's research profile or funding research in developing countries. They believe that more attention should be paid to conventional infrastructure needs in poorer countries which, because of low credit ratings, will have little access to private capital.

The reaction from developing countries will be an important test of the new strategy. The richer countries of southeast Asia, Latin America, North Africa and the Middle East are likely to be more receptive than poorer countries, particularly in sub-Saharan Africa, where the bank is not popular, and where almost 50 per cent of bank-assisted projects have failed during this decade. **Ehsan Masood**

US spirit is willing, but funds are still weak

[WASHINGTON] International collaboration has never been a higher priority for the United States, if public pronouncements by the leaders of the scientific establishment are anything to go by.

Bruce Alberts, president of the National Academy of Sciences, has made collaboration with the developing world one of his top two priorities (the other being science education in schools). "A generous sharing of knowledge resources by our nation's scientists and engineers can improve the lives of those who are most in need around the globe," he told the academy's annual meeting last April.

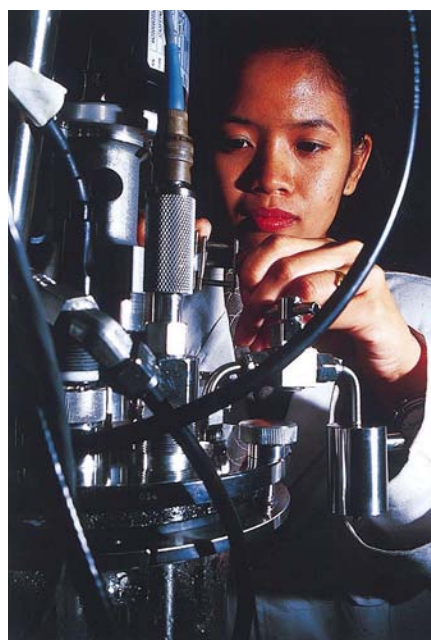
Harold Varmus, director of the National Institutes of Health (NIH), has placed new emphasis in the past year on the need for the NIH to confront global health problems, such as malaria and African strains of AIDS.

And Rita Colwell, the new director of the National Science Foundation (NSF), has spent much of her first few months in office travelling abroad, talking in particular about her collaborative experiences in cholera research (see *Nature* 396, 202; 1998).

But critics say there is more talk about such collaboration than action. And, by raising the issue now, these three scientific leaders are perhaps acknowledging that, in practice, the large and well-funded US scientific community has never been more isolated from the rest of the world.

"We've moved a long way backwards in the past 40 years," says Alberts, adding that a new generation of researchers "has little knowledge about the opportunities and challenges" of working with scientists abroad. "The young people have enormous potential interest" in such research, "but they don't see how to do it," he says.

Of the \$75 billion that the US government will spend on research and development this year — a quarter of it on basic research — only a small fraction will involve collaboration with foreigners. The NIH, the science



Worldwide prescription: the NIH is set to pursue a global public health agenda for the first time.

agency least constrained in spending money abroad, says it will spend \$200 million (1.5 per cent of its budget) on international projects.

The international division of the NSF will spend \$20 million, although other activities at the agency involving foreign partners will cost ten times as much. "They get crumbs from the table," mutters an international official at one leading US scientific society.

Some critics go further in taunting the US track record in supporting international collaboration in science. In the current issue of the journal *Issues in Science and Technology*, Congressman George Brown (Democrat, California) and Daniel Sarewitz of Columbia University's science policy unit write that "although there was little evidence that [international science and technology agreements] have led to significant scientific part-

nerships, there is plenty of evidence that they support a healthy bureaucratic infrastructure in the US government".

Critics also contend that the international division at NSF and the Fogarty International Center, which performs an analogous function at the NIH, have a marginal impact on the powerful directorates and institutes that dominate the two agencies.

One man planning to change that is Gerald Keusch, a biologist with a strong track record in researching infectious diseases in Africa, who was picked by Varmus last September to direct the Fogarty centre. Keusch says Varmus left him in no doubt that the NIH is ready to pursue "a global public health agenda — which is something different from NIH's agenda in the past". In the coming year, the centre's budget will grow by 25 per cent, to \$35 million, although it will remain the smallest budget of any NIH institute.

The Fogarty centre has been shifting its collaborative emphasis from work with scientists in developed countries to work in developing countries and the former Soviet bloc. Between 1987 and 1996, the proportion of its spending on the latter two groups doubled, to more than 40 per cent of the total.

Keusch expects this trend to continue. "My goal is to steadily focus on the developing and transitional countries where needs are greatest," he says. He adds that other mechanisms exist to support collaboration between developed countries.

Keusch is the first director of the centre to also be appointed associate director at the NIH — at the recommendation of an external study into Fogarty's effectiveness. He hopes that this, together with his own scientific relationship with the directors of several powerful NIH institutes, will enable him to increase the centre's influence.

He is also working to improve relations with the World Health Organization, the World Bank and the US Agency for Inter-

national Development, and pledges to track "more effectively" what NIH really spends on international work. The \$200 million, he says, "is only the part that is easily tracked. There are other expenditures that aren't included."

Under Alberts' presidency, the National Academy of Sciences has pursued a phalanx of initiatives to foster collaboration with developing countries. These include a study with China and India into the impact of population growth on land use, to be published in February, and a major conference on sustainability planned for May 2000 in Tokyo.

John Boright, director of international affairs at the academy, says it is also working to bring together groups of young scientists with common interests. But he does not have the money to do this on a significant scale. "We're planting some seeds, but planting the whole field is a different story," Boright says.

Many scientists assume that isolationists in the Congress are the main obstacle to substantial US investment in international science. But congressional staff say the scientific agencies have not asked for more money for international work.

"The agencies use Congress as a bogey-

man," says one staffer, who favours spending on international collaboration. "All the major scientific collaborative programmes began under President Reagan, when money was tight. Now that the money's fine, they don't want to do anything with people overseas."

But some top scientists, at least, beg to differ. "The opportunities for the United States to play a role [in the developing world] are enormous," says Sherwood Rowland, foreign secretary of the academy. "The scientific community is willing to do it, but there isn't sufficient funding. And that reflects the political will of the country." **Colin Macilwain**

Science moves up Europe's aid agenda

[MUNICH] "Five years ago, research was almost a dirty word in development aid circles," says Barend Mons, a senior adviser to the Dutch research organization NWO. Today, however, research is no longer considered "a luxury toy for rich countries", but a fundamental component of economic success in all countries.

The change represents a major shift in European attitudes towards Third World aid in the past few years. Research has been a significant beneficiary.

In June 1997, the European Union's Council of Development Ministers passed a resolution acknowledging the potentially important role of research and technology in achieving the objectives of the EU's development policy. It stressed "the strategic role that research could play in enhancing sustainable development".

This has committed the European Commission, the EU's executive arm, to finding new ways of boosting the amount of research it supports in developing countries — and ensuring that support is effective.

Until relatively recently, most of this support took the form of researcher-initiated projects within the five-year Framework research programmes. The rest was provided through separate development aid programmes designed in collaboration with the governments of recipient countries.

But a review of this strategy in 1997 revealed a lack of synergy between the two arms, and noted that scientists from the South were frequently junior, rather than equal, partners in research projects.

The review also pointed out that research, particularly that funded with development aid money, tended to focus on short-term technical fixes, rather than helping developing countries to increase their self-sufficiency by building up local research capacity.

The underlying problem, according to the commission, had been the lack of political support for research in sustainable development. But many European politicians



Target for aid: Bangladeshis seeking to escape flooding make their way to a cyclone shelter.

now accept that research is a necessity if developing countries are to be helped to solve their problems.

Key phrases now heard in Brussels and increasingly echoed in national development aid policies include 'co-ordination of research effort', 'building research capacity in the South', and 'equal partnerships'.

A small office has recently been established in the commission's development directorate to promote this new philosophy. One model being used is the so-called 'fisheries initiative', a programme set up by the development and research directorates in the early 1990s to address in a coordinated way the problem of diminishing fishing resources in the Third World.

Programme spokeswoman Cornea Noun says the initiative was drawn up after meeting representatives from developing countries, who told the commission what they needed, and what the cultural and political implications of meeting such needs were likely to be.

The programme has a strong research component, and each project it supports has to include partners from the EU and the

South on an equal footing. It must also help build research capacity in the South. "We have practical research tools to develop aquaculture technology," says Noun. "But we must make sure that management tools are in place to allow research-based developments to become general practice."

Coordination of research projects supported by the commission, national development agencies, and other donors such as aid charities, is helped by new web pages such as the health page SHARED (<http://www.shared.de>). This lists and describes all research programmes involving Third World scientists that are supported in Europe, at the national or international level.

One of the new office's goals is to encourage developing countries to spend more of their EU aid on research. At present only one per cent of the EU's ECU3.5 billion (US\$4.1 billion) development budget is spent on research, and the commission would like this to rise to between three and five per cent.

But such growth may take several years to achieve. Lobbies for science are weak in developing countries, whose politicians are

ZED NELSON/PANOS

seldom keen to attend meetings with the commission on research issues.

Yet their support is essential if there is to be any increase in international funding for such work. There is no increase in the money earmarked for the Third World in the EU's new fifth Framework programme of research, even though the ECU250 million handled in the fourth Framework programme is generally considered to have been well spent.

Equal partnerships and the potential for building up local research capacity in the South became mandatory components of all projects funded through the EU's INCO-DC (International Coordination – Developing Countries) programme.

In one project, for example, scientists from France, Britain, Kenya and Brazil collaborated on a study of the genetic component of susceptibility to schistosomiasis. This involved epidemiological studies, mapping of susceptibility genes and the evaluation of immunological correlates of disease.

The study required the creation of an immunology research group at the Faculty of Medicine of Uberaba, Brazil, and strengthening the schistosomiasis research group at the Federal University of Bahia, Brazil.

INCO-DC applies the same rules to its approval of projects as other EU research programmes, emphasizing the need for scientific excellence and eschewing any sense of offering 'charity'. "You don't do anyone any favours by funding second-rate research with second-rate equipment," says a programme spokesman. "Scientists in the South must be supported only to do research at the same standards as in Europe."

EU member states have been adopting the same new philosophy in their national programmes. Denmark, Europe's largest aid donor in relation to its gross national product (GNP), has in the past ten years increased the proportion of its aid budget spent on research nearly fivefold, to three per cent of the total.

One exception to the EU trend is Britain, which devotes only 0.27 per cent of its GNP to development, four times less than Denmark in relative terms, although it spends around seven per cent of this on "research and related knowledge generating activities". Poverty eradication is the main development priority but it "sees continued investment in knowledge generation as a key element in achieving its aims and objectives for international development".

Britain has shown that such philosophical statements can be backed with cash. Last summer, for example, Prime Minister Tony Blair announced to the meeting in Birmingham of the G8 group of industrialized countries that Britain planned to spend £60 million (US\$100 million) to support the international Multilateral Initiative on Malaria (see *Nature* 388, 219; 1997). **Alison Abbott**

Technology and tradition clash in India

[NEW DELHI] When India's first waste incineration plant, designed to generate 4 MW of electricity by burning 300 tonnes of rubbish daily, was commissioned in New Delhi in 1988, officials billed it as the technological solution to the problem created by mountains of municipal waste.

Such plants were to be built nationwide. It was therefore a surprise when the plant closed down in 1992, after remaining idle for four years, and without having produced any electricity (see *Nature* 359, 763; 1992).

The reason was straightforward: neither the government nor the Danish contractor had considered Delhi's 8,000 rag pickers, who systematically retrieve reusable items such as wood, plastics and cloth from municipal landfill sites. Once such combustible materials were removed, the remains – mostly rotten vegetables and unburnable cooking wastes – were insufficient to operate the plant.

"We have learnt that technology by itself cannot solve our society's problems," says Valangiman Ramamurthi, secretary to the Department of Science and Technology (DST). "Science-based innovation has little role in development unless it is socially accepted and fits into the prevailing cultural system."

India's social and cultural diversity also means that what is welcomed by one group may be resented by another. Piped water has been a boon to the people of Tamilnadu, but, according to Ramamurthi, young women in rural Rajasthan found the water taps a threat to rural freedom, and smashed them up.

Before the taps were installed, the girls fetched water from distant sources, and the long daily walks gave them a chance to socialize with men. As well as spoiling their love life the taps added to their drudgery at home, because their mothers, finding that the girls had free time, made them do extra work.

Equally frustrating for Indian researchers has been their years of effort to improve the cycle rickshaw, the primary mode of transport for millions of Indians since its introduction in 1930. A motorized version introduced in the late 1970s never really caught on, as rickshaw pullers became concerned about the need to maintain the motor and the hassles of getting a driving licence.

A similar fate befell a cycle rickshaw fitted with a 'regenerative' braking system (pictured), in which the energy lost in braking was stored in a bank of springs and released when the vehicle was ready to move again. Amitabha Ghosh, the inventor of the device – and now director of the Indian Institute of Technology in Kharagpur – says rickshaws fitted with this system "required 44 per cent less pedalling force during start-ups, besides reducing the energy expended by pullers by 38 per cent".

Despite this advantage, the design has



Inappropriate technology: 'regenerative' brakes on rickshaws ignored the cost of retrofitting

failed to win converts. The main reason, says Ghosh, is that few pullers own the rickshaw – they rent them. "Owners have no motivation for spending \$60 for retrofitting, and have no concern for the health of the pullers, as there is no shortage of poor people in search of an income."

The DST's more recent effort to introduce another versatile design also met with resistance from cycle rickshaw owners in Agra who, fearing competition, smashed up the two vehicles sent for a demonstration.

'Taraloom', an improved hand loom developed with money from the DST, is another example of how a technological innovation, despite its advantages, can fail to have an economic impact unless it is affordable. The rugged steel loom with a flywheel attachment for high-speed operation means weavers can increase output by 60 per cent.

But the 15,000 rupees (US\$354) price is beyond the reach of millions of poor weavers who work at home on wooden hand looms passed down through generations. In the past eight years, only 20,000 Taralooms have been sold, mostly to cooperatives. "Cost is just one factor," admits Sanjay Sharma, inventor of Taraloom. "The important barrier is tradition."

'Modernity' is also being put to the test at Agra, where the Taj Mahal's white marble has turned yellowish due to sulphur dioxide pollution from the 170 coal-fired foundries nearby. "Twenty-five years ago we offered to replace their traditional furnaces with gas-fired ones but they refused," says P. Ramachandra Rao, director of the National Metallurgical Laboratory (NML) in Jamshedpur.

Now the foundries have been ordered to adopt NML technology by the supreme court. Rao says that the \$30,000 cost of a gas-fired cupola is the main reason for the foundries' opposition. But commercial sources admit that the foundries are resisting natural gas because of the reduced opportunities for tax evasion. Unlike coal, natural gas is supplied by the government and metered. Production figures can be worked out from gas consumed, and tax levied.

K. S. Jayaraman

Internet may help bridge the gap

[PARIS] Until recently, many feared that increasing use of the Internet would widen yet further the knowledge gap between industrialized and developing countries. But such pessimism is giving way to optimism that the new technology may instead help spur a renaissance of science and technology in poorer countries.

Indeed, the prospect that Internet access will eventually become widely available seems to be shifting the focus of the debate onto issues related to access to and the production of the content of websites.

Use of the Internet in developing countries is now growing faster than anywhere else. According to a report from the International Data Corporation, the number of users in the Asia-Pacific region will grow from 6.5 million last year to 29.3 million by 2001, while the number in Latin America, Africa and Eastern and Central Europe will rise from 7.6 million to 25.6 million.

Lack of access to the Internet has been the largest barrier to its use in developing countries. "But this is changing extremely fast," says Ben Fouché, a South African who is vice-president of the International Council for Scientific and Technical Information, which

was set up by the International Council for Science (ICSU) to promote the availability of scientific information worldwide.

Ironically, while developing countries have long struggled with inadequate copper-based telecommunications infrastructures, this lack is now proving an advantage. Whereas advanced countries are reluctant to replace their entire telephone systems, many developing countries, such as Gambia, are installing digital networks using the latest fibre-optic technology.

"Developing countries have the chance to leapfrog [advanced countries] because the infrastructure is becoming more affordable," says Fouché. Prices are being driven down by the liberalization of telephone monopolies, and by commercial operators looking to developing countries as future markets.

Joshua Rosenthal, from the Fogarty International Center of the US National Institutes of Health, points out that whereas development of the Internet in the North was initially driven by the research community, in the South scientists will piggyback on commercial investment.

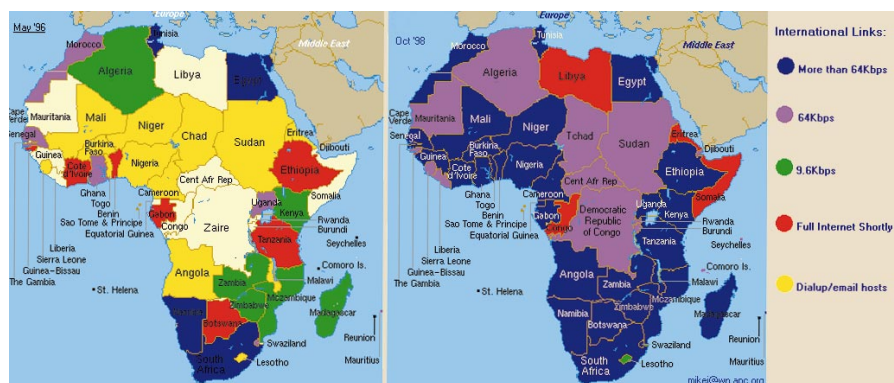
One of the most ambitious of many commercial investments is Africa One, a proposal

to throw a 39,000-km optical-fibre submarine cable around Africa. The US\$1.6-billion plan was launched by the International Telecommunications Union, with the support of 30 African countries, and companies including British Telecom and AT&T.

A plethora of initiatives is under way to boost Internet networks for universities in developing countries, for example at the World Bank, the United Nations Educational, Scientific and Cultural Organization and other UN bodies — the last mainly through an informal group of agencies known as PICTA, Partnership for Information and Communication Technologies in Africa (www.bellanet.org/partners/picta).

Widespread access to the Internet allows scientists to use a vast range of tools, literature and databases. It promises to offset the drop in journal acquisitions by libraries in developing countries that has occurred over the past decade because of inflation.

"Less developed countries stand to benefit at least equally [to Northern scientists] from technological developments," says Paul Ginsparg, founder of the Los Alamos physics e-print archives, which has set up mirror sites in India, China, Russia and Brazil.



Quick change: how electronic communication in Africa grew between May 1996 and October 1998.

But there is resentment among scientists in developing countries at what they see as a lack of effort by publishers in the North to use the lower costs of distribution on the web to make online versions of journals available to them at lower prices than print versions.

Fouché points to the “paradox” that, although telecommunications companies are prepared to discount physical infrastructure in view of a future market, publishers seem reluctant to follow this example, even though the costs would be much less.

“Publishers have lacked a strategy for developing countries in terms of long-term market development. They are not asking what should we be doing to make informa-

tion content affordable now.”

One of the few organizations with a pricing policy for developing countries is HighWire Press, a not-for-profit body set up by Stanford University to help universities and learned societies publish on the web at low cost (hwmg.stanford.edu/developing.html).

HighWire has arranged with the publishers of many of the 100 or so leading journals it puts on the web for reduced subscriptions for countries with a per capita gross national product of US\$9,000 or less. Such countries can often receive the journals for as little as one-third of the US price of print versions. The UK Institute of Physics also offers free access to its 33 electronic titles

to academic institutions in Africa.

The American Physical Society is considering national online-only site licences for countries in Eastern Europe and the former Soviet Union. Academic Press does not have specific licences with developing countries, but is “open to discussion”, according to Chrysanne Lowe, its marketing director.

Scientists in developing countries are increasingly taking advantage of the web to publish their own journals. Rosenthal says that by doing so they will be better able to communicate between themselves, and to overcome some of the hurdles and prejudice that have hampered their efforts to publish in the journals of industrialized countries.

The Fogarty centre has launched a programme with the US National Library of Medicine to train scientists from developing countries in information technology (www.nih.gov/fic/opportunities/itmi.html). Two bodies have also been set up to help scientific societies publish online — the International Network for the Availability of Scientific Publications (www.oneworld.org/inasp), established by ICSU, and the UK Electronic Publishing Trust for Development (dialspace.dial.pipex.com/town/street/fu84), set up by bioscience publisher Bioline Publications and a Brazilian bioinformatics body. Between them, they have launched two dozen web journals.

Declan Butler

UK societies lobby against cut in school science curriculum

[LONDON] Ten of Britain's foremost professional science bodies have challenged the government to end speculation that it is thinking of reducing the time pupils spend studying science in the last two compulsory years of secondary school.

The challenge comes in a joint statement issued last week by a group of organizations that included the Royal Society, the Association for Science Education, and the Royal Academy of Engineering.

Britain's national curriculum, set up ten years ago, is currently being reviewed. One possible outcome is a reduction in the 20 per cent of curriculum time that most pupils aged between 14 and 16 spend studying science. Nigel Thomas, education manager at the Royal Society, says the statement is aimed at dissuading the government from adopting such a course of action.

Minister pledges to boost Russian science spend

[MOSCOW] Valentina Matvienko, Russia's vice prime minister, promised last month that the government will "substantially increase" its financing of science and social needs in 1999. Speaking at a meeting in Moscow on 22 December, she said that the budget recently submitted to the State Duma (the lower house of the Russian parliament) proposed spending 80.6 billion rubles (US\$3.8 billion) on science and social needs.

This combined total represents an increase of 21.8 per cent on the amount formally allotted in 1998 — and is 73.9 per cent more than was actually spent. Fundamental science and new technologies receive 11.4 billion rubles, 2 per cent more than in 1998. But this amount is still only 2.0 per cent of the budget expenditure, whereas federal legislation requires that 4 per cent is spent on science.

Indian scientists told to become trend-setters

[CHENNAI] Politics took precedence over science in the agenda of India's prime minister Atal Behari Vajpayee on 3 January when, after hurriedly inaugurating the Indian Science Congress at Chennai (formerly Madras), he immediately left to keep political engagements at Bangalore, failing to attend a traditional reception with delegates.

In a short speech, Vajpayee asked Indian scientists to become "trend-setters" focusing on high-quality work that would stand up to international scrutiny. He urged them to

work closely with the best institutions in the world, but added that they should be prepared to stand on their own feet "if any country tries to deny us the opportunities of legitimate scientific cooperation".

Industrial research still soars upwards in US

[WASHINGTON] Research and development expenditure by private industry in the United States is set to surge by more than 9 per cent this year to a total of \$157 billion, according to an annual survey conducted by Battelle and *R&D Magazine*. The survey, which predicts industrial research trends by asking corporations about their plans and extrapolating the results to cover the entire US economy, says the federal government's spending on research and development will increase by only 2 per cent, to \$68 billion.

Because of heavy industrial investment, total US R&D spending as a percentage of gross domestic product rose from 2.4 per cent in 1994 to 2.6 per cent in 1998, according to government figures. The federal government's share of that — which was more than half as recently as 1980 — has now fallen to less than 30 per cent of the total. Industry's investment is mainly in applied research and development, however, and the government still supports most basic scientific research.

South Africa ministry official stays on

[CAPE TOWN] Roger Jardine, who announced he was quitting as director-general of South Africa's Department of Arts, Culture, Science and Technology with effect from 31 December last year after a clash with his minister, Lionel Mtshali, has agreed to remain in office until May (see *Nature* 396, 298; 1998).

Jardine is thought to have been persuaded to stay on until the election, due in May, which is certain to be followed by a cabinet reshuffle. The presidential review commission has proposed combining the department's functions with those of others after the election.

Solar observatory flies into yet more trouble

[WASHINGTON] Scientific observations with the recently recovered Solar and Heliospheric Observatory (SOHO) have been suspended after the last of three onboard gyros apparently failed late last month (see *Nature* 396, 399; 1998).

The gyro was used only occasionally, but was necessary for normal operation of SOHO's guidance system. Without it, the European-US spacecraft uses fuel at an excessive rate, according to Joseph Gurman

of NASA's Goddard Space Flight Center, the US project scientist for SOHO.

Project managers hope to begin a temporary mode of operation that will allow most scientific observations to resume by the end of this month. Meanwhile, the European Space Agency is revising the spacecraft's guidance software, a permanent solution to the problem that could take several months.

Sex discrimination suit at Lawrence Livermore

[SAN FRANCISCO] Six women are filing a class-action lawsuit claiming gender discrimination on behalf of more than 3,000 female workers at Lawrence Livermore National Laboratory in California.

The women allege that they have been systematically underpaid and under-promoted for decades at the lab, and say that they earn between \$1,000 and \$2,000 less per month than male counterparts. Their attorneys say back pay could average \$250,000 each. If the courts find in their favour, the laboratory could be forced to reimburse all the women it has employed. The laboratory is managed by the University of California's Board of Regents.

Spacecraft heads off to take a dig at Mars

[WASHINGTON] The Mars Polar Lander blasted off from Cape Canaveral in Florida on Sunday (3 January), beginning its 11-month journey to become the first spacecraft to land near the edge of the planet's southern polar cap.

The lander is equipped with a robotic arm, which will dig one metre into the martian surface. Two microprobes will then conduct two days of soil and water experiments. The spacecraft is the second to be launched to Mars within three weeks.

Alas, poor scientists, Shakespeare wins poll

[LONDON] Charles Darwin and Isaac Newton have failed to win the title of greatest Briton of the millennium in a national poll. More than 11,000 people chose William Shakespeare as 'Personality of the Millennium' from an all-male short-list drawn up by the BBC from votes cast in a poll of radio listeners.

Darwin and Newton polled fourth and fifth place, with even their combined votes failing to match those of Shakespeare. Prominent living scientists have been quick to express their dismay and question the high positions of other individuals short-listed, such as Winston Churchill and William Caxton, England's first printer. Michael Faraday and Alexander Fleming just failed to be short-listed.

AIDS decision threatens drugs trials

Sir—Your editorial, “Setting a bad example on AIDS”, correctly points out the immediate consequences of the decision by South Africa’s health minister, Nkosazana Zuma, not to fund the supply of the drug AZT for HIV-infected pregnant women (*Nature* 396, 603; 1998). But the impact goes far beyond the ultimate death of possibly thousands of infants who will be denied the potential benefit of AZT.

There are far-reaching implications which threaten the often difficult and complex proceedings that determine the design of clinical studies in developing countries; the commitment of pharmaceutical companies which evaluate therapies for developing countries; and the willingness of such companies to negotiate reduced prices for developing countries.

Current perinatal HIV prevention studies, aimed at reducing cost while preserving efficacy, have been designed with developing countries in mind. The availability of a drug following completion of a study in the host country is part of the ethical deliberations that are seriously considered before the approval of a clinical

study. While an absolute guarantee is not required, the ethical review committee requires a certain degree of assurance that the host country and the pharmaceutical company will act in good faith to try to make a drug available.

Zuma’s decision has escaped no one’s attention. South Africa receives many grants from foreign agencies for clinical drug studies and hosts many drug studies sponsored by pharmaceutical companies. Pharmaceutical companies entered into negotiations with several developing countries to supply AZT at reduced cost to prevent perinatal HIV transmission. Zuma’s decision may affect the willingness of scientists, funding agencies and companies to embrace trials in South Africa or other developing countries.

In addition, health ministers of poorer countries might use South Africa as an example to justify suspension of other public health measures, and pharmaceutical companies may begin to view drug evaluation in developing countries as too risky. The decision threatens to undermine the efforts of many

of us who are trying to convince companies that they must provide approved drugs for evaluation in developing countries and support investigations of new drugs. Alienation of the academic community and pharmaceutical companies threatens to ‘bite the hand’ that has the potential to control the HIV epidemic.

It is not for scientists or funding agencies in developed countries to dictate what must be done in developing countries, nor for pharmaceutical companies to determine what drugs are necessarily best for a country with limited resources. However, if scientists, funding agencies and pharmaceutical companies respond to the desperate needs of developing countries and receive assurances that agreed commitments will be honoured, it is to no one’s benefit to withdraw these commitments when there is a consensus that this will result in the unnecessary loss of thousands of lives.

Arthur J. Ammann

*Global Strategies for HIV Prevention,
104 Dominican Drive, San Rafael,
California 94901, USA*

Money talks louder than research quality

Sir—There would be few who would disagree with Gerard Vassiliou’s comments about the exploitation of young scientists by ambitious group leaders in large ‘factory’ laboratories (*Nature* 396, 307; 1998). As a newcomer myself (PhD 1995) I certainly appreciate the problem. One of the most obvious reflections of this depressing trend is the almost total absence of single-author papers in leading journals.

Nowadays it seems one cannot do important research without joining a large group. What was not emphasized, however, was that this problem is an inevitable symptom of the academic reward system, which recognizes not just the quality and quantity of research, but also the associated funding. So researchers are hired and promoted partly on the operating budget and size of their group, and universities are ranked on the amount of research funds they attract. If two scientists in the same field produce work of equal quality and quantity, they should be rated identically. But the current reward system does not do this—the researcher who attracts more money is preferred.

The reward system is now so skewed that many people in less expensive fields (such as

my area, animal morphology) feel compelled to apply for grants, not because they need the money to do research, but because they need to attract money to satisfy the powers that be.

Michael Lee

*Department of Zoology, University of Queensland,
St Lucia, Brisbane, Queensland 4072, Australia*

Lysenko took a dismal view of Malthus

Sir—The rejection of modern genetics by Stalin’s biologist Trofim Lysenko is well known. Less well known is that he also rejected Thomas Malthus’s insights into population biology which are so central to Darwinism. Roger Short¹ notes the hostility of Marx and Engels to the “Dismal Theorem” of Malthus in which human population increases exponentially, subsistence increases arithmetically, and misery results. Lysenko went further and threw the baby out with the bathwater.

On 31 July 1948, Lysenko gave his opening address to the now notorious meeting of the Lenin Academy of Agricultural Sciences². His first target was Malthus: “Many are still not clear about Darwin’s error in transferring into his teaching Malthus’ preposterous reactionary

ideas on population.... Darwin himself was unable to fight free of the theoretical errors of which he was guilty. Today there is absolutely no justification for accepting the erroneous aspects of the Darwinian theory, those based on Malthus’ theory of overpopulation with the inference of a struggle going on within species. And it is all the more inadmissible to represent those erroneous aspects as the cornerstone of Darwinism.”

What must it have felt like to hear the cornerstone of Darwinism dismissed as an error? It was the “classics of Marxism” that revealed this and other errors in Darwinism to Lysenko: “Biologists should always ponder these words of Engels, ‘The entire Darwinian teaching on the struggle for existence merely transfers from society to the realm of living nature Hobbes’ teaching on *bellum omnium contra omnes* and the bourgeois economic teaching on competition, along with Malthus’ population theory.... The childishness of this procedure is obvious, and it is not worth while wasting words on it.”

Damning words, indeed. But five days later, one of Lysenko’s fiercest opponents, B. M. Zavadovsky, launched a scathing attack on his intellectual impostures. He pointed out that this quotation was from a private letter Engels had written to a friend and he analysed the relevant sections of true Marxist classics: Engels’ “Anti-Duhring” and the “Theory of Surplus Value” by Marx.

The opinions of Marx and Engels are revealed as being rather more subtle. They accepted that human populations have the capacity for exponential increase, hence their delight that Darwin had disproved the “Dismal Theorem” by observing that ‘subsistence’ (that is, animals and plants) also has the capacity for exponential increase. Marx writes: “Darwin failed in his excellent work to see the fact that by discovering ‘geometrical’ progression in the animal and plant world, he was refuting the theory of Malthus. The Malthusian theory is based precisely on the point that he counterposes the geometrical progression of man to the ‘arithmetical’ progression of animals and plants.”

Not only was Lysenko not a Darwinian, he was not even a Marxist, according to the analysis of Zavadovsky.

On 7 August 1948, Lysenko began his closing address by informing the conference that he had the endorsement of the Central Committee of the Communist Party for his views, and a letter that morning in *Pravda* confirmed this news. The debate was over. A third-rate scientist had risen to prominence through the backing of a ruthless dictator: how fortunate that such things no longer occur!

Sean Nee

Department of Zoology, University of Oxford,
South Parks Road, Oxford OX1 3PS, UK

1. Short, R. *Nature* 395, 456 (1998).

2. Opening address by Academician T. D. Lysenko in *Proc. Lenin Acad. Agric. Sci. USSR*, July 31–August 7, 1948 (Foreign Languages Publishing House, Moscow, 1949).

Beauty and the Bart Simpson effect

Sir—Douglas Yu and Glenn Shepard Jr have shown that, although it is widespread, the phenomenon that men find pictures of women more attractive if the image has a low waist-to-hip ratio is not culturally invariant¹.

This aesthetic does not extend to Matsigenka men, isolated deep in Manu Park, Peru. These men prefer thick-waisted women. A good thing, since that apparently describes healthy young Matsigenka women. The authors make the reasonable suggestion that the apparent invariance of preferences in earlier studies “may have only reflected the pervasiveness of western media”. Apparently, Matsigenka men do not share this aesthetic because “their degree of isolation is about as high as can be obtained today”.

I found this result so interesting that I prepared a lecture on it for my behaviour class. I searched the World-Wide Web for pictures of the Matsigenka (also spelled ‘Machiguenga’) and found something ironic and amusing. At a site maintained by E. Russo, who collaborates with Shepard,

there is a photo of a Matsigenka couple, Mateo and Aleja². Mateo is holding an ocelot. Aleja is wearing a Bart Simpson T-shirt. The Matsigenka’s isolation may be “about as high as can be obtained today”, but they are not isolated enough to escape Bart Simpson. Luckily, it is not a Marge Simpson T-shirt. That would have suggested that the preference for thick-waisted women might also have been influenced by western media.

Thomas Getty

Kellogg Biological Station, Hickory Corners,
Michigan 49060, USA

1. Yu, D. W. & Shepard, G. H. Jr *Nature* 396, 321–322 (1998).

2. <http://www.montana.com/manu/village.html>

English versus Spanish in science evaluation

Sir—Gianmarco Paris *et al.*¹ consider how the frequency with which an article is cited is affected by its country of origin. They base their information on the Science Citation Index (SCI), an Institute of Scientific Information service comprising approximately 3,500 of the world’s leading scientific and technical journals covering a broad range of disciplines. A parallel problem concerning this excellent database is seriously affecting some European and Latin American countries². This is the misuse of the SCI in relation to language of publication in certain fields of work that are characterized by territoriality (for example, Earth science).

Major problems with its use can occur when: (1) publication in SCI journals is used as the only criterion for evaluating the scientific productivity of researchers and when it is used to confer prestige, stipends and even promotion; and (2) scientific tribunals pay more attention to the impact factor of the journal than to the quality of the scientific contribution.

Bearing in mind the difficulty of comparing quite similar CVs, the SCI-based evaluation provides an easier judgement procedure. But it is unfair to scientists from countries whose journals are poorly or not represented in the SCI. A good example of this biased analysis has recently been pointed out by the Centre for Scientific Information and Documentation (CINDOC) of the Spanish Consejo Superior de Investigaciones Científicas (CSIC). The CINDOC is responsible for producing and distributing the ICYT database, a set of contributions in science and technology that has covered 500 Spanish bibliographic sources at an average annual rate of 6,500 items since 1979.

The CINDOC study covers a five-year period, surveying the productivity of the whole Spanish community of Earth scientists, and obtaining the opinion — using a specially created, 27-point survey —

of 383 of these researchers³. The design of the study allows it to indicate the extent of the language problem, as no Spanish-language Earth science journal is currently included in this SCI category.

The results obtained indicate that, whereas 73% of these scientists usually publish in both SCI and Spanish journals, most of their articles (69%) are published in domestic journals. Therefore, the application of a solely SCI-based evaluation does not properly reflect their total scientific productivity and thus makes them unhappy with this evaluation procedure (58.5% of the whole, and more than 70% of researchers from universities and CSIC). When other databases besides the SCI and ICYT are included so as to give a wider range of Earth science journals (for example, GeoRef), the Spanish contribution to Earth science is found to be 7.3% of European production, occupying sixth place after the United Kingdom, France, Germany, Holland and Italy. The corresponding number derived from the SCI database is 5.6%.

It is true that the lack of publication in SCI journals can indicate the low quality of some scientific contributions and/or the authors’ inability to publish in high-quality international journals. But it is also true that scientific authorities should not disregard the idiosyncratic nature of some research fields, or the difficulty of publishing domestic work in international journals that demand topics be of “general interest”. Thus, a frequent answer of journal editors is: “the paper is good but it would fit better in a local or regional journal”. They apparently fail to realize that, due to the preponderance of English in science, many English papers are just as domestic as the Spanish ones. However, English-language journals that publish domestic science, such as *J. Geol. Soc., Lond.*, get into the SCI, whereas the equivalent Spanish-language journals, such as *Rev. Soc. Geol. Esp.*, do not.

If Europe wants a strongly united scientific community, it should define evaluation criteria that do not undervalue the quality of a publication simply because it is written in the ‘wrong’ language.

Jesús Rey-Rocha & M. José Martín-Sempere

Centre for Scientific Information and Documentation (CINDOC), Consejo Superior de Investigaciones Científicas (CSIC), Joaquín Costa, 22. 28002 Madrid, Spain

Fernando Lopez-Vera

Department of Agricultural Chemistry, Geology and Geochemistry, Faculty of Sciences, Universidad Autónoma de Madrid, Spain

Jesus Martinez-Frias

Departamento de Geología Museo Nacional de Ciencias Naturales (CSIC), c/ José Gutiérrez Abascal, 2 28006 Madrid, Spain

1. Paris, G. *et al. Nature* 396, 210 (1998).

2. García-Guinea, J. *Nature* 379, 109 (1996).

3. Rey-Rocha, J. Thesis. Univ. Autónoma Madrid (1998).

Eighteen-ninety-nine and all that

This year's occasions for anniversary celebrations include the invention of the Schnelltelegraph and aspirin, the unearthing of the Rosetta stone, many features of zinc, the first numerate pope, and information glut.

J. L. Heilbron and W. F. Bynum

The discovery that makes the anniversary of the year was the battery, invented by Alessandro Volta in 1799 and made public, through a letter to Sir Joseph Banks, in 1800. The world-wide celebrations, which began last year and will continue into the next, make the sixth set of international commemorations of Volta. Orators have made him an anticlerical hero of the unification of Italy (he was a devout Catholic), a bridge between the Roman and Fascist empires (he was neither imperious nor imperial), and, after the Second World War, the exemplar of correct behaviour in defeat (he allowed himself to be patronized by Napoleon after the French conquest of Lombardy).

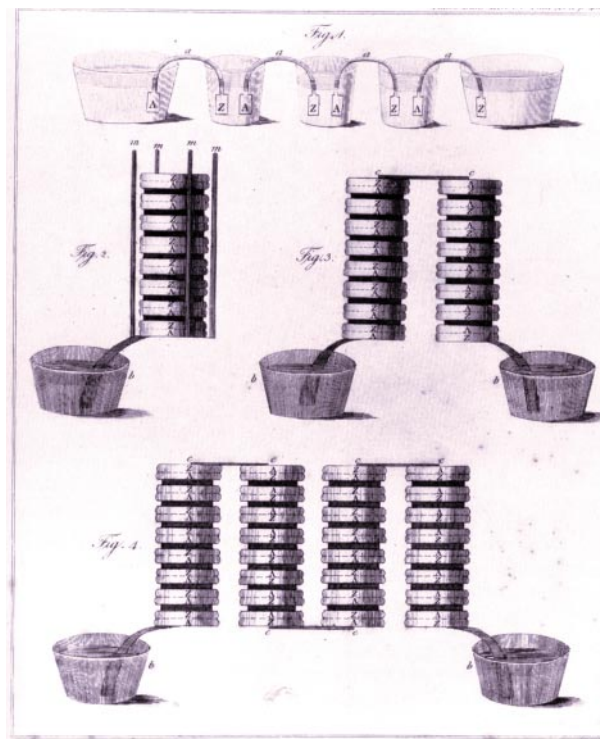
A better theme for this year of the euro is Volta as a European. He was at home in Paris and Geneva, visited Göttingen and London, and announced the discovery of the battery, the consequence of a creative competition between Italians, in a letter written in French to the Royal Society of London.

We begin as usual with eligible centennials, then sesquicentennials, and so on, back 15 centuries, to 499, before returning to our own time and to the anniversaries of outstanding events that happened 50 or 50 ± 25 years ago. As usual, € signifies a centennial and \$ indicates a Nobel prize.

1899 (1.0€): Decadence and decay

In Germany, 1899 was a typical *fin-de-siècle* year. Georg Cantor and David Hilbert broke their heads worrying whether the paradoxes of the set of all ordinal numbers would wreck the theory of sets. Ernst Haeckel fretted over the *Welträtsel* — the riddle of the Universe; his readers fretted over his conclusion that the mind does not survive the body. In its mercy, providence gave the Farbenfabriken Bayer aspirin, now 100 years old. The defining drug of our century, it would not today pass regulatory hurdles. Further to the relief of pain, cocaine was used as a spinal anaesthetic, by August Bier (died 1949, 0.5€), and novocaine was synthesized by Alfred Einhorn.

The prolific discoveries in physics that began with the discovery of X-rays in 1895 continued to yield decays. Following up the discovery of radioactivity, André-Louis Debierne (died 1949, 0.5€), a friend of the Curies, found actinium. Marie Curie (\$ physics 1903), Henri Becquerel (\$ physics 1903), and Friedrich Oskar Giesel independently showed that the beta rays Ernest Rutherford (\$ chemistry 1908) had identi-



Stacks of new ideas: in 1799 Alessandro Volta showed that dissimilar metals in salt water generate an electric current.

fied in uranium rays could be deflected in a magnetic field. And Rutherford himself, and also Friedrich Ernst Dorn, detected the emanation from thorium. These discoveries soon gave rise to entire research fields: actinium, like thorium and uranium, turned out to have its own radioactive decay series. The deflectable beta rays turned out to be streams of fast electrons useful for probing atoms and testing relativity. And thorium emanation turned out to be the clue to the uncovering of the natural transformations of the elements.

Jacques Loeb (born 1859, 1.4€) liked to show that certain newly hatched caterpillars, which crawl up to the tips of branches to eat buds, could be made to starve to death by shining a light at them from below. This *fin-de-siècle* behaviour, he said, showed that the caterpillars did not run to the buds through some mysterious instinct but as a mere mechanical response to light. Loeb followed up these experiments, which date from around 1889 (1.1€), with efforts to reduce the equally mysterious business of procreation to plain physics and chemistry. Those who believe, as he did, that life is a mechanical process should delight in commemorating his demonstration, a century ago, that, when properly treated chemically, sea urchins' eggs can be made to reproduce parthenogenetically.

In his delicious book, *Degeneration* (1895), Max Nordau (born 1849, 1.5€) listed

the afflictions that would probably kill civilization in the twentieth century. They included information glut. Since he was a pessimist, he was no doubt pleased in 1899 by the invention of the Schnelltelegraph by Josef Virág and Antal Pollák of Budapest. By 1902 the system could take down 100,000 words per hour, in Hungarian. The octuplex system devised by Henry Augustus Rowland, of the grating, was less menacing. As a professor of physics, Rowland decried his colleagues who commercialized their discoveries; but, as a man dying of diabetes, he did what he could to provide for his family. The octuplex of 1899 handled four telegrams in each direction simultaneously.

The *fin de siècle* gave us metal that does not expand and wood that does not burn. The first, invar, an alloy of iron and nickel, was the invention of Charles Edouard Guillaume (\$ physics 1920), long-time director of the Bureau International des Poids et Mesures. The firm of Hölsberg in Frankfurt am Main introduced wood made incombustible by impregnation with a solution of boric acid and ammonium sulphate. It, or something like it, is still sold world-wide as firewood.

We record the deaths of the German chemist Robert Wilhelm Bunsen, of the burner, a pioneer of spectroscopy and other methods of inorganic analysis; Sir Edward Frankland, another chemist, and also a

humanitarian, concerned to improve the quality of drinking water; the Norwegian mathematician Marius Sophus Lie, eponymous creator of the Lie group; and Othniel Charles Marsh, American palaeontologist, known especially for his classifications of dinosaurs, early horses and other extinct vertebrates.

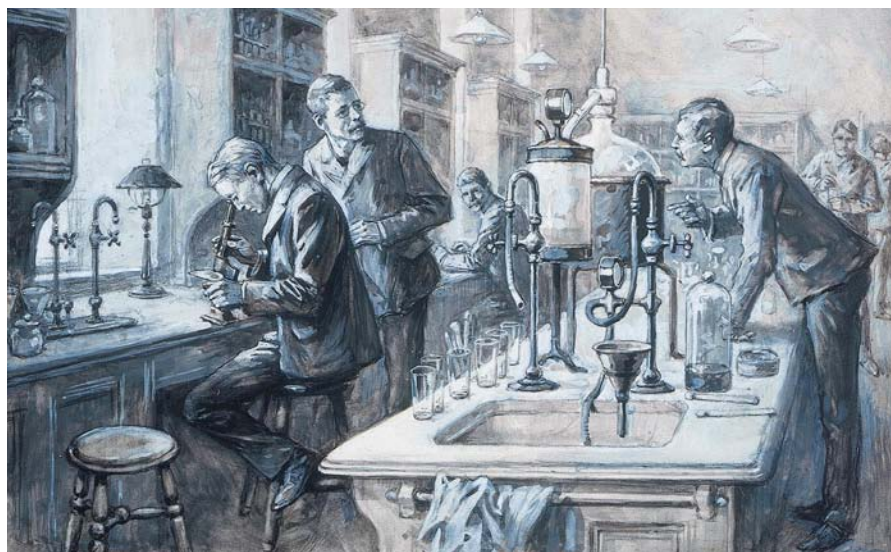
These extinctions were balanced by the births of Paul Hermann Müller, the Swiss chemist who declared war on mosquitoes with DDT (\$ medicine/physiology 1948); of Max Theiler, the South African virologist who produced an effective vaccine against a mosquito-borne disease, yellow fever (\$ medicine/physiology 1951); and of the London and (by some reckonings) the Liverpool Schools of Tropical Medicine.

1849 (1.5¢): Warm ice and sick zinc

The hot news in physics in 1849 was James Thomson's deduction from Carnot's principle that freezing should occur at higher temperature with increasing pressure. Conversely, water created under pressure applied to ice should refreeze when the pressure relaxes. And so it was and is. This same year James's brother William, later Lord Kelvin, introduced the term 'thermodynamics' to designate the field in which they were interested.

Zinc buffs have a once-in-a-millennium opportunity this year. The Edward Frankland whose death we reported earlier (see 1899) discovered zinc alkyl and other metallo-organic compounds, and Gabriel Auguste Daubrée crystallized zinc oxide, both in 1849; whereas, we now recall, in 1899 (1.0¢) the Dutch chemist Ernst Julius Cohen (born 1869, 1.3¢) identified three allotropic forms of zinc. Thus 'zinc disease', which had been known since the time of Aristotle, received an explanation: zinc darkens at normal temperatures owing to a transition from the white form to the grey form around 20 °C.

Viagra takers may be interested in the fact that 150 years ago A. A. Berthold showed that



Feverish activity: a laboratory in the Liverpool School of Tropical Medicine, founded in 1899.

testes implanted in castrated cocks prevent symptoms of castration. One who may have needed Viagra was Thomas Addison, a physician at Guy's Hospital in London, who identified pernicious anaemia in 1849, two years after beginning a childless marriage to a woman with two children. Aloys Pollender, a physician in a small town in Germany, presents a different case. Twenty years after he claimed to have discovered the anthrax bacillus (in 1849, of course), he fell in love with a woman 42 years his junior. He married, for the first time, at the age of 70, fathered a son, and lasted nine years in wedlock. He died in 1879 (1.2¢).

Pollender married for love and to avoid scandal. Claude Bernard married without love to obtain the financial security to pursue his scientific career, which in 1849 was thriving. He brought to fruition a decade's work on sugar metabolism and the role of the pancreas in digestion, and discovered how to produce sugar in the urine of his experimental dogs through selective brain ablation. His wife darkly suspected that the dogs had died in vain.

We mention the deaths of Johann Wolfgang Döbereiner, a chemist best known for his 'triads', relationships among atomic weights regarded as forerunners of the periodic table of the elements, and of Sir Marc Isambard Brunel (born 1769, 2.3¢), who drove the first tunnel under the Thames and sired Isambard Kingdom Brunel (died 1869, 1.3¢), the designer of the *Great Eastern* steamship, from which the first successful Atlantic cables were laid; and the births of Luther Burbank, prolific American hybridizer of fruits and vegetables, and Sir William Osler, that most admired of modern physicians.

1799 (2.0¢): Dead and kicking

Volta made the anniversary-of-the-year invention by following up the odd discovery

by Luigi Galvani that a freshly dissected frog could be made to kick by connecting a leg nerve to its corresponding muscle. The oddity was that the connector had to consist of two different sorts of metal joined together seriatim. Galvani inferred that the agent causing the kick was a special sort of fluid, an 'animal electricity', concocted in the frog's brain and stored in its muscles much as ordinary electricity was generated by a machine and warehoused in Leyden jars. As an anatomist, Galvani focused on the animal's parts and functions. As an electrician, Volta zeroed in on the bimetallic conductor. He came to place the generation of the electricity responsible for the kick in the contact of dissimilar materials.

To meet Galvani's resourceful challenges to this interpretation, Volta had to eliminate the frog, which served not only as the generator of animal electricity (in Galvani's view) but also as its detector (in everybody's experiments). Volta refined the standard electrical measuring devices, several of which he had invented, to detect a contact potential of less than one volt. Then he amplified the effect by taking dozens of silver coins and an equal number of discs of our theme metal zinc (see 1849), putting them together in pairs, and stacking the pairs with bits of card soaked in salt water between them. In devising the artifice of the soaked card spacers he probably had in mind the watery environment and repetitive structure of the special organs of the torpedo, an electrical fish whose performance had been shown by Henry Cavendish to depend on ordinary electricity.

While Volta was making the future, the savants accompanying Napoleon's expeditionary force to Egypt were digging up the past. In 1799 they found the Rosetta stone, which can claim notice in *Nature* because the first person to use it effectively in deciphering hieroglyphics was Thomas Young, of the modulus, the reinventer of the wave theory



Deep designs: the first tunnel under the Thames, built by Marc Isambard Brunel, who died in 1849.

of light. To keep his savants busy, Napoleon set up an Institut d'Égypte in weak imitation of the then-new Institut de France. He made the military engineer and mathematician Gaspard Monge its president. While Monge presided abroad, his wife deceived him with his chief disciple at home. No, not that way. They published his *Application de l'analyse à la géométrie*, a book he seemed unable to finish, which established the field of descriptive geometry.

On seizing power by the coup of 18 Brumaire (9 November) 1799, Napoleon appointed the mathematician Pierre-Simon Laplace (born 1749, 2.5¢) his minister of the interior. Laplace could not grasp the big problems. Napoleon observed that he carried the spirit of the infinitesimal into administration and sacked him in six weeks. The first consul may have dipped into the first two volumes of his minister's *Mécanique céleste*, then just published, which Laplace had given him. If so, he might have noticed the proof that the Solar System is stable under gravitational forces. That might have prompted him to ask, as the story has it, what role God, an essential cog in Newton's system, played in the world, and Laplace to reply, "Sire, I have no need of that hypothesis". It was no doubt a bold claim since, as far as Laplace knew, no one had yet even proved the fundamental theorem of algebra. In fact, Carl Friedrich Gauss had just supplied the proof (the first of four he was to find) in his doctoral thesis of 1799. Perhaps we should have mentioned in its place that the world had to wait until 1899 for David Hilbert, in his *Grundlagen der Geometrie*, to clean up the mess left by Euclid.

We notice the deaths of Joseph Black, a physician who became professor of chemistry at the University of Edinburgh, the discoverer of 'fixed air' (carbon dioxide), the first of the gases that powered the chemical revolution, and the inventor of the concepts of specific and latent heats; the Dutch Catholic Jan Ingen-Housz, physician and electrician, who inoculated the children of the Holy Roman Empress and discovered photosynthesis; the Genevan Horace Bénédict de Saussure, one of the first men to reach the summit of Mont Blanc, where he spent four hours making measurements with instruments carted up by his 18 guides; the English physician William Withering, member of the Lunar Society of Birmingham, who had learnt of the power of digitalis from an old woman in Shropshire; and Lazzaro Spallanzani (born 1729, 2.7¢), Volta's prolific and ornery colleague in natural history at the University of Pavia, who showed that decapitated snails could regrow their heads, and also investigated the vascular system, infusoria, digestion, reproduction, volcanology, and much more.

Among promising beginnings there were Alexander von Humboldt's departure for his



Not today, Josephine: Napoleon's camp-followers discovered the Rosetta stone in 1799.

famous five-year trip to the New World; William Smith's (born 1769, 2.3¢, died 1839, 1.6¢) private communication of his work on the relation between fossils and strata; and the births of John Lindley, English botanist, horticulturist, publicist and wordsmith (we owe him 'conifer' and 'orchid'), who championed natural systems of classification over the artificial sex scheme of Linnaeus, and of the Royal Institution of London.

1749 (2.5¢): Labour of love

To begin with a beginning, Georges-Louis Leclerc, Comte de Buffon, published the first of the 36 volumes of his influential *Histoire naturelle*. It too begins at the beginning, with the formation of the Earth by collision of a comet with the Sun 75,000 years ago. The same year, 1749, Franklin proposed that someone should show the identity of lightning and electricity by tapping a thunder



Died reckoning: the Marquise du Châtelet translated Newton before her early demise.

cloud. Buffon saw the merit of the proposal and of having someone else implement it. He hired a retired dragoon, who, with the help of a *curé*, performed Franklin's experiment in 1752. Fortunately for them, they caught electrical disturbances of the lower atmosphere, not a lightning bolt, on their probe.

Our death of the year features Gabrielle-Émilie le Tonnelier de Breteuil, Marquise du Châtelet. The marquise, a close student of the works of Leibniz and Newton, and also of her lover Voltaire, undertook around 1745 to translate Newton's formidable *Principia mathematica philosophiae naturalis* into French. She had almost finished early in 1749 when she noticed she was pregnant, the result of preferring the attentions of a young officer to the delights of mathematics. She redoubled her efforts, fearing, all too correctly, that her confinement would leave her little time for Newton. She died of childbed fever. Her manuscript, published ten years later (1759, 2.4¢), is still the only French translation of the *Principia*.

1699 (3.0¢): Nothing of note

We know that in 1699 Guillaume Amontons observed that air increases in volume by about one third when heated from the freezing to the boiling point of water, but we do not think it worthy of celebration.

1649 (3.5¢): Christian Epicure

For 1649 we can offer Pierre Gassendi's *Animadversiones... de vita, moribus, placitisque Epicuri*. Gassendi was a professor of mathematics in Paris, an astronomer famous for his observation of a transit of Mercury, a cosmologist inclined towards Galileo, and, what then took agility, a priest convinced of atomism. His book christianizes Epicurus via a nominalist or fictionalist epistemology, much as the Jesuits tried to adapt Coperni-

cus. The atomic theory as redeployed by Gassendi, together with the system of whirlpools of his contemporary and sparing partner René Descartes, marked out the via moderna during the scientific revolution.

1549 (4.5¢): An unlikely story

Our sources say that in this year two barber surgeons of Naples, Zocarello and Fioravanti, cured a woman suffering from dropsy by removing her spleen, but we doubt it.

1449 (5.5¢): Martyr to mathematics

Ulugh Beg, 'the Great prince', the peace-loving grandson of the fierce Tamerlane, founded a school emphasizing astronomy in Samarkand in 1420. Four years later he built an observatory incorporating a huge sextant, 40 metres in radius, fixed in the plane of the meridian. It delivered a good value for the obliquity of the ecliptic, only half a minute out. The school prepared trigonometric and astronomical tables and an important catalogue of stars. Ulugh Beg died in 1449, violently. His observatory was destroyed. Not everyone likes mathematics.

999 (10¢): Numero uno

The millennial award goes to Gerbert d'Aurillac, as much of a mathematician as it was possible to be in a society ignorant of Euclid. It is not for his arithmetic, however — which ended in the construction of an abacus with 27 columns, or his approximations to the square roots of two (17/12) and three (26/15) — that we mention him, but for becoming pope, the first mathematician to do so, 1,000 years ago.

499 (15¢): Fruit of desire

Five hundred years before Gerbert became pope, Aryabhata I summarized what the Hindus knew of mathematics in a book called the *Aryabhatiya* (cf. *Berkeleyum*, 1949). Among matters of perhaps greater importance, he gave π as 3.14156, derived from consideration of a regular polygon with 384 sides. If more teachers wrote as he did, mathematics might be more popular. Here is an example: "In the rule of three, multiply the fruit by the desire and divide by the measure. The result will be the fruit of the desire."

1949 (0.5¢): Scary stuff

The heavens are full of portents for those who can read them. In the year of NATO, the Soviet atomic bomb, and the foundation of the two Germanys, Wilhelm Heinrich Walter Baade, German-American astronomer best known for his investigations of distant stars, discovered the asteroid Icarus, a menacing Earth-grazer, which has come within four million miles of us. Almost simultaneously, and closer, the United States flew its first two-stage rocket.

While we are among the stars, we should mention the conjecture, by Fred Lawrence



Mothering molecules: Dorothy Crowfoot Hodgkin away from the electronic computer.

Whipple, that even the most glorious of comets is nothing but a dirty snowball. Snowballs remind us that the Nobel prize for chemistry in 1949 went to William Francis Giauque of the University of California at Berkeley for his technique of adiabatic demagnetization and other cryogenic work. It was a good year for his modest university. Its Radiation Laboratory announced its creation of elements 98 and 99, to which it shamelessly gave the names Californium and Berkeleyum.

The other Nobel prizewinners in science that year were Hideki Yukawa in physics and, in physiology or medicine, Walter Rudolf Hess and Antonio Caetano de Abreu Freire Egas Moniz. Yukawa, the first world-class Japanese physicist educated entirely in Japan, showed the value of his education by publishing his prizewinning theory as his first paper, three years before earning his PhD. Moniz won his for orientating psychiatry in an organic direction. Postulating that psychoses result from synaptic fixation, he had the happy idea of encouraging his unhappy patients into new thought patterns by severing their frontal lobes.

Under old business we can mention the announcement by Linus Pauling (\$\$ chemistry 1954, peace 1962) that one class of the disease recognized by Thomas Addison (see 1849), sickle-cell anaemia, arises from an abnormality in the haemoglobin molecule caused by a defective gene. Under new business there is an *embarras de richesse*: Dorothy Crowfoot Hodgkin (\$ chemistry 1964) initiated the use of the electronic computer to find the structure of an organic molecule (it was penicillin); Claude Elwood Shannon introduced the world to his information theory; the BIVAC, the first American stored-program computer, started up; Harold Lyons built the first atomic clock; and Peter

Brian Medawar (\$ physiology/medicine 1960) showed that his technique of introducing foreign proteins into mouse embryos could be applied against the rejection of transplants.

1924, 1974 (50 ± 25¢): Take your partner

This year we introduce an innovation in our art. We select among the discoveries and inventions that occurred 25 or 75 years ago only those that can be combined in pairs to indicate a half century's progress or regress. For progress, take space exploration. We regard 1924, when the English translation of Hermann Oberth's *Die Rakete zu den Planetenraumen* appeared, as the average date of publication of the most influential work of the rocket pioneers Oberth, Konstantin Tsiolkovsky and Robert Goddard. They were prophetic, but did not get beyond the stratosphere. Fifty years later a Soviet probe landed on Mars and Mariner 10 flew by Venus and Mercury.

For regress take anthropology. In 1924 Raymond Dart discovered *Australopithecus africanus*, which, at between 100,000 and 300,000 years of age, was regarded as the oldest known humanoid. In 1974, Donald Johanson and Maurice Taieb dug up 40 per cent of Lucy and reckoned her age at three million years. Both groups of workers had the good sense to report their findings in *Nature*.

A few further examples might be useful. In 1924, Hans Spemann (\$ physiology/medicine 1935) and his student Hilde Mangold showed by moving bits of embryos around that one part can influence the development of another; in 1974, 139 members of the US National Academy of Sciences called for a one-year moratorium on research in genetic engineering. In 1924, Edward Victor Appleton (\$ physics 1947) sounded the ionosphere by radio waves and discovered the layers named after him; in 1974, the Nobel prize for physics went jointly to Martin Ryle and Anthony Hewish of Cambridge University for their contributions to radio astrophysics, including the discovery of pulsars.

To bring this year's budget of anniversaries to a close, we observe that in 1924 Harry Steenbock created vitamin D with the help of ultraviolet light, which gave rise to a beneficent industry to improve food by UV irradiation, and that in 1974 F. Sherwood Rowland and Mario Molina warned that chlorofluorocarbons destroy the ozone layer, which gave rise to a fear that we will receive more UV light than is good for us. We can't resist one more: 1924 saw the first marketing of the spiral-bound notebook and 1974 ± 2 that of the Hewlett-Packard programmable pocket calculator. □

J. L. Heilbron is at Worcester College, Oxford OX1 2HB, UK. W. F. Bynum is at the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BE, UK.

Mother and father in surprise genetic agreement

Mark Pagel

In genomic imprinting, one copy of a gene is switched off depending on whether it came from the father or mother. One explanation centres on the competing interests of the two parents. Another now brings in sex-associated differences in the offspring.

Even detractors from the idea that genes are selected to maximize their own survival must grudgingly acknowledge beauty in the phenomenon of genomic imprinting. Many genes vary their effects depending upon the environment in which they find themselves; for example, genes for antlers or long canine teeth are frequently silent in females. Other genes spread to the detriment of their hosts. But imprinted genes are altogether more subtle: an imprinted gene behaves differently depending on which parent contributed it to the offspring. It knows whence it came, and the same gene acts one way if contributed by the father and another way if contributed by the mother.

The leading theory of imprinting proposes that mothers and fathers have conflicting interests in how much the mother should invest in their offspring: fathers wish to have mothers make large babies, mothers wish to spare some resources for future offspring^{1,2}. Most known imprinted genes are involved with embryonic growth and fit neatly with this theory. Steadily, however, evidence accumulates of imprinted genes that seem to have nothing to do with embryonic or other growth patterns. A theory proposed by Iwasa³, based on the principle of 'dosage compensation', now provides an explanation for these observations. It opens up a new theoretical domain to explain why imprinted genes may even affect social skills in boys and girls.

Offspring of sexually reproducing individuals have two copies of nearly all of their genes, one from the mother, the other from the father. Normally both copies are expressed. But imprinted genes carry some sort of a message — the gene may get tagged by methylation^{4,5} — which switches off one of the copies. This same tag makes it possible to identify, and for the gene to behave as if it knows, its parentage.

A well-studied case is that of the *Igf-2* gene, which produces an insulin-like growth factor that operates during embryonic development. When this gene is expressed, demands are placed on the mother to pro-

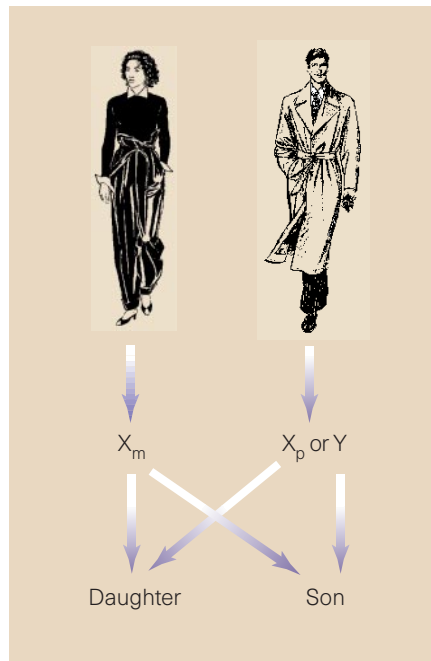


Figure 1 A father's X chromosome always contributes to a daughter's genetic make-up, the mother's to a son's or a daughter's, meaning that the paternal X chromosome can contribute daughter-specific information on sex differences. The theory proposed by Iwasa^{3,8,9} is that genomic imprinting may have evolved as a mechanism of dosage compensation to silence the maternal X or the paternal X as appropriate. (Modified from ref. 9.)

vide more nutrients to the embryo. In mice and humans, copies received from the mother are switched off, those from the father are not. It is as if the otherwise identical maternal and paternal copies of the gene have adopted opposite and competing strategies that act to cancel each other in the developing embryo.

To explain this phenomenon, Moore and Haig¹ proposed that during embryonic growth it might be in the paternally derived copy's interest to demand more of the mother than the maternally derived copy. The reason? If the mother is likely to produce children by more than one father, her offspring

will be less related to each other than if the mother was monogamous. A baby that under these circumstances places extreme demands on its mother may disadvantage her future offspring. But these future offspring will be step-siblings and thus less likely to share genes than full siblings. Confronted with this tactic, the maternal copy's best strategy is to switch off. Monogamy is rare in animals, even in humans. Of the 1,154 human societies documented in the *Ethnographic Atlas*⁶, 1,073 practise some form of polygyny in which females may have children by more than one male. Genomic conflict is one cost of infidelity.

Other embryonic growth genes show similar imprinting effects, and the conflict theory has received broad support. But how can the conflict theory explain the observations of the powerful role that genomic imprinting plays in Turner's syndrome⁷? The X and Y chromosomes are the sex-determining chromosomes in mammals. Girls normally receive two X chromosomes, one from each parent. Boys receive an X from their mother and a Y chromosome from their father. Turner's syndrome arises when a girl receives only one X chromosome. If it comes from the mother, the girl, among other things, is likely to be socially disruptive, behaviour that most parents will unhesitatingly recognize as normal in little boys. If she receives her X chromosomes from the father, however, her behaviour is closer to normal for that of a girl. Little girls indeed are Daddy's little girls.

It turns out that the X-chromosome genes associated with Turner's syndrome are imprinted. The conflict theory, so successful with growth genes, cannot easily accommodate this. The theory derived its force from the competing interests of mothers and fathers. But, independently of how many different males a female has children by, there is no necessary conflict between mothers and fathers over sex differences in behaviour. If what is good for a boy differs from what is good for a girl, then mothers and fathers will both wish to see these differences made manifest.

Iwasa's dosage-compensation explanation (ref. 3; see also refs 8 and 9) recognizes that, because females have two X chromosomes and males only one, females potentially receive twice the dose of gene products found on the X. This can be disruptive to other genes shared by males and females, and so females have evolved to switch off genes on one of their X chromosomes, and imprinting appears to be one way they do it.

If females, for whatever reasons, have been selected to have greater social skills than males, then that information must be contributed by the paternal X. This is because the paternal X chromosome is always contributed to a daughter, but maternal X's can go to sons or daughters (Fig. 1). Under

these circumstances imprinting can evolve to provide a ready and precise mechanism for dosage compensation and, in this case, for paternal expression.

The same logic can be applied to understand the production of other sex differences by X-imprinted genes. In mice, female embryos that lack one of the X chromosomes differ in size depending upon whether the missing chromosome is the paternal or maternal copy: those lacking the paternal copy are larger¹⁰. This is opposite to the pattern expected from the conflict theory. In normal mice, females are smaller than males. If the growth genes that produce this sex difference are carried on the X chromosome, then imprinting the paternally contributed copy to code for smaller size will produce the advantageous sex difference. If it was the maternal copy that carried this information, daughters might be small, but so might be sons.

Despite setbacks, the conflict theory of imprinting can coexist peacefully with the dosage compensation/sex differences theory, as they explain different evolutionary settings. Nevertheless, battles are certain to be fought over which theory applies in any given context. In its favour, genomic-imprinting-based dosage compensation is a powerful and very general force that might be used whenever it is advantageous for sons and daughters to differ, be it in behaviour, morphology or physiology. It also does not depend upon the existence or degree of polygyny, or any other factor that might pit mothers' and fathers' interests against each other.

Agreement between mothers and fathers, surprising as it may seem to evolutionary biologists brought up on conflict, may prove to extend even beyond dosage compensation

and sex differences. There is striking evidence that imprinting is involved in mammalian brain development, with the maternal copies of genes promoting growth of the cortex, while paternal copies retard it^{11,12}: the maternal genome may be responsible for at least part of the rapid expansion of the mammalian brain over evolutionary time.

No one knows why this pattern of gene expression has evolved. One possibility is that, in a social species, it is advantageous for sons and daughters to inherit their mothers' brand or style of intelligence — after all, by virtue of surviving to reproduce she has shown that her cognitive style is suited to her environment. Mothers and fathers can, in evolutionary terms, agree on this, and use imprinting to fine-tune genetic expression. In the meantime, while these intriguing new phenomena are being investigated further, fathers can at least take heart that it is they who are to be credited for their daughters' exemplary social behaviour. □

Mark Pagel is in the School of Animal and Microbial Sciences, University of Reading, Whiteknights, Reading RG6 6AJ, UK.
e-mail: m.pagel@reading.ac.uk

1. Moore, T. & Haig, D. *Trends Genet.* **7**, 45–49 (1991).
2. Haig, D. & Graham, C. *Cell* **64**, 1045–1046 (1991).
3. Iwasa, Y. *Curr. Top. Dev. Biol.* **40**, 255–293 (1998).
4. Ferguson-Smith, A. C. *et al. Nature* **351**, 667–670 (1991).
5. Birger, Y., Shemer, R., Perk, J. & Razin, A. *Nature* **397**, 84–88 (1999).
6. Murdock, G. P. *Ethnographic Atlas* (Pittsburgh Univ. Press, 1967).
7. Skuse, D. H. *et al. Nature* **387**, 705–708 (1997).
8. Iwasa, Y., Mochizuki, A. & Takeda, Y. *Evol. Ecol. Res.* (in the press).
9. Iwasa, Y. & Pomiankowski, A. *J. Theor. Biol.* (in the press).
10. Thornhill, A. R. & Burgoyne, P. S. *Development* **118**, 171–174 (1993).
11. Keverne, E. B., Martel, F. L. & Nevison, C. M. *Proc. R. Soc. Lond. B* **262**, 689–696 (1996).
12. Keverne, E. B. *et al. Dev. Brain Res.* **92**, 91–100 (1996).

The static structure and equilibrium framework has yielded fundamental insights and led to success in designing drugs, yet we probably know more about the protein-folding reaction coordinate than the protein–ligand reaction coordinate. For example, looking at mechanism and dynamics, there is a much clearer picture of folding intermediates and transition states than for corresponding pathways by which ligands exit their binding sites. It is fair to ask whether there are even well-defined ligand-exit pathways that can be described by a reaction-coordinate model. We do not know whether ligands dissociate through a random collection of different pathways and barrier topologies, or whether they dissociate through a well-ordered sequence of bond-breaking and conformational events. Most importantly, how can we get at these fundamental questions experimentally?

We can look at the structure and energetics of transition states indirectly by transition-state mapping, an approach that has been used to characterize enzyme and protein-folding reaction coordinates. But force microscopy^{3,4} now offers a striking new approach for mapping energy-barrier topologies as ligands are pulled away from their binding sites. To study protein–ligand interactions, a probe is modified with the ligand and the protein immobilized on one of a variety of surfaces (Fig. 1). The probe/ligand are brought into binding contact with the protein, then the force required to disengage them is measured. However, interpreting this force measurement has been difficult and controversial because many parameters could potentially contribute to it, including non-specific interactions between the probe and surface. Moreover, the measured force typically represents an unknown number of protein–ligand interactions that vary in their spatial arrangements owing to the geometry of the probe surface (which is much larger than the individual proteins and ligands) and variable linker orientations.

What, then, does the force measurement represent? Merkel *et al.*¹ provide a convincing theoretical framework, backed by experiment, showing that the strength of protein–ligand bonds varies with loading rate. They experimentally isolated single streptavidin–biotin interactions (the prototype of earlier receptor–ligand force studies). Then, by measuring interaction survival times and strengths over loading rates spanning six orders of magnitude, they mapped the prominent barriers along the force-driven activation landscape. Atomic force microscopy (AFM) studies⁵ first showed that force techniques could probe specific biomolecular interactions, and attempts were made to relate the measured strengths to equilibrium⁶ or activation⁷ thermody-

Biophysics

May the force be with you

Patrick S. Stayton

Many of us grew up with the simplifying assumptions of biology in a test tube — dilute solutions, purified biomolecules and the like. But the biomolecular machinery evolved in a much more complex environment than this. I can remember, several years ago, seeing a cartoon that depicted the remarkably crowded and organized environment of the cytoplasm. I was struck for the first time by the implications of chaperonins for the protein-folding field. It isn't that the biophysics we learned in the test tube is wrong, or cannot be extrapolated to the living cell, but we must acknowledge that questions of relevance and adequacy sometimes arise.

So too for the field that deals with the molecular details of how proteins and their ligands recognize each other. It is a field

dominated by a static viewpoint, where interactions are measured in the test tube and receptor–ligand energetics are characterized by equilibrium thermodynamics. But on page 50 of this issue, Merkel *et al.*¹ report that they have used force measurements on single molecules to show that protein–ligand interaction strengths and dynamics are not fixed in stone. Because the receptor–ligand interaction is governed by thermal activation, the interaction strength and dynamics can be tuned to be weak or strong by varying how fast the retracting force is driven (the 'loading rate'). Nature is a wonderful engineer and, in cases where variable mechanical loading is biologically relevant², it might regulate bioadhesive strengths through activation barriers engineered by evolution.



Figure 1 The streptavidin–biotin receptor–ligand pair. Merkel *et al.*¹ attached streptavidin to a surface and biotin to a probe. Having allowed the two to interact, the authors then measured the forces needed to pull them apart. They found that the interaction strengths depend on how fast the ligand is pulled away from the protein — they seem weak or strong depending on how quickly or slowly the retracting force is driven.

dynamic parameters. Merkel *et al.* now show that these AFM measurements represent only a single point in a continuous ‘dynamic’ spectrum of the force-driven activation-barrier landscape. In a striking proof with the high-affinity streptavidin–biotin pair, which has a dissociation half-life on the order of days in solution, they show that individual interactions can survive for as little as 0.001 seconds, with a strength of only 5 pN at slow loading rates. The interaction strengths increase to around 170 pN at high loading rates, consistent with values measured in previous AFM experiments.

Can we relate this force-driven energy landscape to the molecular coordinates of the streptavidin–biotin dissociation pathway? Fortunately, three different lines of investigation have converged which, taken together, indicate this might be possible. First are the force-microscopy studies. Second, molecular dynamics simulations of the unbinding mechanism under force have been done with both streptavidin⁸ and avidin⁹. The unbinding activation landscape obtained by experiment can be qualitatively mapped onto these simulations, and Merkel *et al.* suggest that the outer activation barrier in their work could correspond to the opening of the flexible binding loop in the protein,

which seems to act as a swinging door for ligand entry and exit. Finally, transition-state mapping of the streptavidin–biotin dissociation pathway has also addressed the energetics and structures along the free (non-force-driven) activation pathway¹⁰. These studies indicate that there is a relatively well-defined exit pathway which looks remarkably like that derived from the force experiments and simulations. This is, perhaps, not surprising given that biotin may exit ‘tail’ first in a similar way to the force-driven pathway, which is dictated by the tether point between biotin’s tail and the surface of the probe (Fig. 1).

We now seem to have an unprecedented opportunity to detail the molecular structure and energetic relationships governing the streptavidin–biotin reaction coordinate by a combination of computational and experimental techniques. But we do still need to know how much of the streptavidin–biotin model is truly ‘model’, and to document how nature and bioengineers can use this analogue-like receptor–ligand bio-

physics. As John Locke noted, “That which is static ... is boring. That which is dynamic and random is confusing. In-between lies art.” □

Patrick S. Stayton is in the Department of Bioengineering, Box 352125, University of Washington, Seattle, Washington 98195, USA.

e-mail: stayton@bioeng.washington.edu

1. Merkel, R., Nassoy, P., Leung, A., Ritchie, K. & Evans, E. *Nature* **397**, 50–53 (1999).
2. Alon, R., Hammer, D. A. & Springer, T. A. *Nature* **374**, 539–542 (1995).
3. Binnig, G., Quate, C. F. & Gerber, C. H. *Phys. Rev. Lett.* **56**, 930–933 (1986).
4. Drake, B. *et al. Science* **243**, 1586–1589 (1989).
5. Lee, G. U., Kidwell, D. A. & Colton, R. J. *Langmuir* **10**, 354–357 (1994).
6. Moy, V. T., Florin, E.-L. & Gaub, H. E. *Science* **264**, 257–259 (1994).
7. Chilkoti, A., Boland, T., Ratner, B. D. & Stayton, P. S. *Biophys. J.* **69**, 2125–2130 (1995).
8. Grubmüller, H., Heymann, B. & Tavan, P. *Science* **271**, 997–999 (1996).
9. Izrailev, S., Stepaniants, S., Balsera, M., Oono, Y. & Schulten, K. *Biophys. J.* **72**, 1568–1581 (1997).
10. Chilkoti, A. & Stayton, P. S. *J. Am. Chem. Soc.* **117**, 10622–10628 (1995).

Ecology

A shrike for mobility

Peter D. Moore

Getting from one place to another is a problem for a plant. One of the most efficient solutions is to wrap seeds in an attractive, preferably tasty, envelope and so encourage gluttonous animals to transport them. Wandering frugivores may thus become inadvertent vehicles of plant migration, but they could also find themselves the target of higher predators — so what then is the fate of the fruits? Is this where the story ends for the passively mobile plant? Apparently not. Work on fruit-eating lizards from the Canary Islands by Nogales, Delgado and Medina¹, reported in the *Journal of Ecology*, shows that viable seeds can be ingested by the predator with its prey and taken to further destinations, thus adding a new dimension to dispersal.

The effectiveness of seed dispersal by passing through animal guts (endozoochory) is widely recognized, and illustrations abound. Telegraph poles and wires over open grassy plains can be the first stage in colonization by trees, as their avian seed vectors find somewhere to perch². Migratory birds may carry plants considerable distances, as in the case of snow buntings (*Plectrophenax nivalis*) collected on the island of Surtsey off Iceland. The gizzards of these birds contained many seeds, including those of the bog rosemary (*Andromeda polifolia*), which must have been consumed in Britain³. So, defecation or death of seed-eating migratory birds can result in colonization and spread of plants to new regions, depending

on how long the seeds have been in the gut, their final germination potential, and the speed and direction of bird movement.

Even flightless birds can be valuable vectors — the Australian cycad genus *Macrozamia*, for example, is dispersed by emus⁴. But flight offers rapid and distant transport that may be particularly important for colonizing islands. For instance, redevelopment of vegetation on the Krakatau Islands between Java and Sumatra after the destructive volcanic eruption of 1883 may have been largely due to the movements of fruit-eating birds and bats⁵. More than 20 species of fig (*Ficus* spp.) must have arrived by air in the guts of flying vertebrates.

Frugivory is less common among reptiles than in birds. However, it is found in many island lizard species including *Gallotia atlantica*, an endemic species of the Canary Islands. This is the only lizard on the small island of Alegranza (10.5 km²), 17 km north of Lanzarote, and it is the focus of the study by Nogales and colleagues¹. There is only one fleshy-fruited plant on the island, *Lycium intricatum* (Solanaceae), a thorny shrub up to 2 m in height that bears bright red berries in winter. The lizard eats many of these berries, and the authors found that about 30% of lizard droppings contained the seeds. The lizard, in turn, falls prey to the great grey shrike (*Lanius excubitor*). Seeds within the lizard’s alimentary system are consumed by shrikes, but they are then often regurgitated in pellets. The fact that the seeds in the pellets

are ingested with lizards — rather than directly from the plant — is supported by observation and by a strong correlation between lizard remains and *Lycium* seeds in the pellets. The seeds in the shrike pellets had a higher germination rate (64%) than those from lizard droppings (50%) or directly from the plant (54%), showing that their experience in passing through two vectors had increased their potential for immediate germination.

Lizards are an effective dispersal system for *Lycium* on Alegranza, but movement of the plant between islands using the lizard as a vector is likely to be rare (although not impossible⁵). Movement of the seeds from lizard to bird, however, presents a range of opportunities for island-dwelling plants, giving them a far greater chance of moving to new islands. Because the seeds are retained in the shrike's gizzard for less than an hour, feeding and flight must follow in rapid succession if dispersal by this means is to be effective. Like so many other events in island biology, given enough time it will probably happen.

The significance of predators as secondary dispersers of fruits and seeds is likely to be limited to specific situations. In Britain,

the sparrowhawk (*Accipiter nisus*) is known to exploit situations where flocks of fruit-eating birds are feeding⁶. So, again, the predator could easily ingest seeds with the prey. But the transfer of seeds from one bird to another in this way is unlikely to have much of an effect on seed dispersal. On Alegranza, *Lycium* seeds have also been found in kestrel (*Falco tinnunculus*) pellets, but these may have been derived from fruit-eating birds rather than lizards. Only where seed transfer involves increased mobility and range (as in the plant–lizard–bird sequence) will new dimensions of dispersal be opened. Predation can then assume a biogeographically significant role. □

Peter D. Moore is in the Division of Life Sciences, King's College, Campden Hill Road, London W8 7AH, UK.

e-mail: peter.moore@kcl.ac.uk

1. Nogales, M., Delgado, J. D. & Medina, F. M. *J. Ecol.* **86**, 866–871 (1998).
2. van der Pijl, L. *Principles of Dispersal in Higher Plants* (Springer, Berlin, 1982).
3. Fridriksson, S. *Surtsey* (Wiley, New York, 1975).
4. Whittaker, R. J. & Jones, S. H. *J. Biogeogr.* **21**, 245–258 (1994).
5. Cendisky, E. J., Hodge, K. & Dudley, J. *Nature* **395**, 556 (1998).
6. Snow, B. & Snow, D. *Birds and Berries* (T. & A. D. Poyser, Calton, 1988).

Geochemistry

Colloidal culprits in contamination

Bruce D. Honeyman

Perhaps nowhere have the details of contaminant transport in groundwater systems been more contentious than in the area of nuclear waste disposal. Over the past decade, the discovery of colloidal forms of actinides, such as plutonium (Pu), has often been at the centre of concern over underground storage of radionuclides. On page 56 of this issue¹ Kersting *et al.* provide an illustration of the striking influence colloids² may have on contaminant transport. The authors have studied groundwater migration of Pu from a nuclear detonation site in Nevada. However, the particular significance of their report lies in reinforcing a general awareness of colloid-facilitated contaminant transport.

As late as the early 1980s, a groundwater contaminant was generally assumed to occur in one of two phases: a stationary phase consisting of aquifer solids; and a mobile, aqueous phase that serves as the medium for the movement of dissolved chemical species. In such a system, the rate of contaminant transport depends on the groundwater transport velocity, of course, but also on the distribution of the contaminant between the two phases. The greater the extent to which a contaminant partitions into the immobile phase, the slower its average transport velocity in the ground water (Fig. 1).

The unexpected appearance of low-

solubility contaminants some distance from known sources, or sooner than would be expected from their solubility, led to examination of the possible involvement of non-aqueous, mobile colloids in contaminant transport³. Invoking colloids to explain such observations gave rise to three-phase models of contaminant transport (Fig. 2, overleaf). The results of Kersting *et al.* reinforce the need for such models. In their work, they use isotopic 'fingerprinting' to demonstrate that Pu, an element with extremely low aqueous solubility (as low as 10^{-17} M, depending on

its oxidation state), was transported significant distances through a groundwater system, and they argue that colloidal species were responsible.

Colloid-facilitated transport of contaminants has become a sort of Gordian knot for environmental scientists. Field studies have often been subject to sampling artefacts, such as colloid mobilization through excessive well-pumping rates, and it has been difficult to reconcile field observations with theory and model simulations. (In 1988, a reviewer of a manuscript⁴, on which I was an author, implicating a colloidal intermediate in the scavenging of metals in marine systems, complained: "The problem developing in the literature with colloids is that they are blamed or claimed for everything that can't be explained".) Since then, colloid-facilitated contaminant transport has gained broad acceptance. Nonetheless, few studies have unequivocally demonstrated its significance in the field.

Figure 2 compares two- and three-phase contaminant transport models. Figure 2a shows a low-solubility contaminant distributed between an aqueous phase and immobile aquifer solids (the macroparticles). Colloidal materials can increase the apparent solubility of low-solubility contaminants (Fig. 2b) if those contaminants strongly associate with the colloids and the colloids minimally interact with the stationary macroparticle phase. Because colloid groundwater concentrations are typically quite low, the contaminants most likely to be transported by colloidal materials are of extremely low solubility and strongly partition to mobile non-aqueous phase materials. In this respect, Pu is an ideal candidate; others include pesticides (for instance, DDT and dieldrin), polynuclear aromatic hydrocarbons, other light actinides (for example, thorium) and many transition metals (such as cobalt).

Groundwater colloids originate from two sources⁵: through mobilization of existing colloid-sized (1 nm to 1 µm) materials in aquifer systems following chemical pertur-

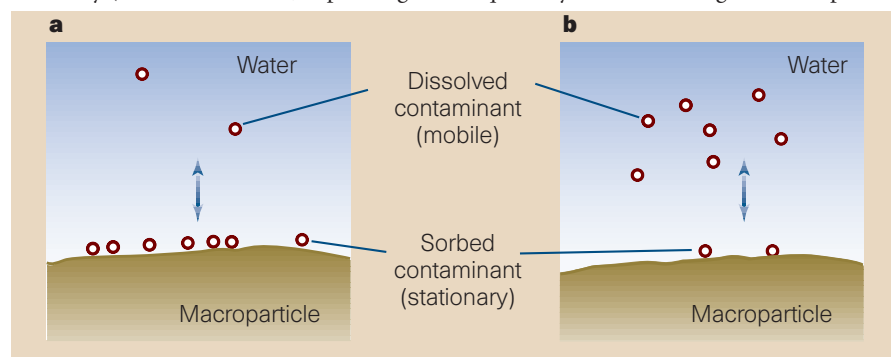


Figure 1 Contaminant transport in a simple two-phase groundwater system. a, High sorption, and therefore low contaminant solubility; b, low sorption, high contaminant solubility. Macroparticles are the stationary components of a groundwater aquifer and include clays, metal oxides such as quartz, and particulate organic matter. The extent to which contaminant molecules are sorpted to macroparticles (by adsorption, surface precipitation or absorption) regulates the rate of contaminant transfer through groundwater systems for a given water velocity.



100 YEARS AGO

The native arithmetic of Murray Island, Torres Strait, is described by the Rev. A. E. Hunt in the latest *Journal* (New Series, vol. I Nos. 1 and 2) of the Anthropological Institute. The only native numerals are netat (one) and neis (two). Higher numbers would be described either by reduplication, as neis netat, literally, two-one for three; neis-i-neis, or two-two for four, &c., or by reference to some part of the body. By the latter method a total of thirty-one could be counted. The counting commenced at the little finger of the left-hand, thence counting the digits, wrist, elbow, armpit, shoulder, hollow above the clavicle, thorax and thence in reverse order down the right arm, ending with little finger or right hand. This gives twenty-one. The toes are then resorted to, and these give ten more. Beyond this number the term gaire (many) would be used. English numerals are now in general use in the Islands. From *Nature* 5 January 1899.

50 YEARS AGO

Fertilized mouse ova have been cultivated *in vitro*, and their development filmed, by Friedrich-Freska and Kuhl, who used as medium a clot of guinea pig plasma and mouse embryo extract containing segments of Fallopian tube. Like Chang, I have been working on ovum culture with a view to transplantation of ova. Chang has used rabbits, with serum as a culture medium; I have chosen to use mice, as more readily available, and because (like most domestic animals) they have naked eggs. Seeking a medium readily prepared in large quantities, I have tried saline hen-egg extracts. The procedure adopted revealed an unanticipated difference between the viabilities of two-cell and later tubal stages; eight-cell ova survived and developed in culture, whereas two-cell ova did not. ... A few became blastocysts either completely separated from the zona, or spherical and still half-enclosed in the split and distended membrane. ... A physiological difference between two-cell and eight-cell stages seems clearly established; the reason for the difference remains obscure. The result recalls a similar one of Chang's, namely, survival of morulae, but not two-cell ova, when transplanted into the rabbit uterus, though both survived when cultivated in serum. From *Nature* 1 January 1949.

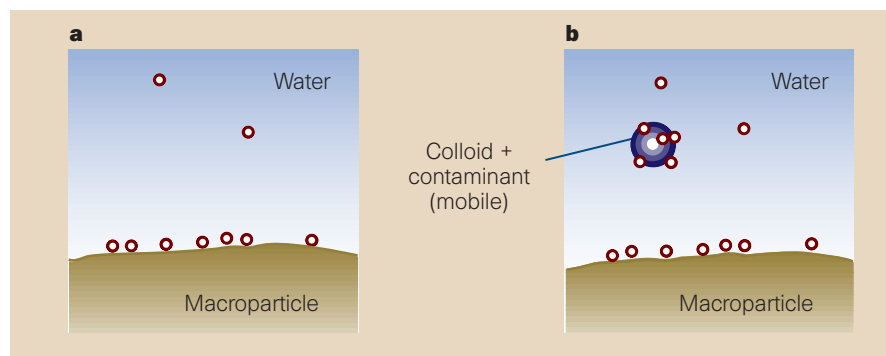


Figure 2 Comparison of generalized two- and three-phase groundwater systems. a, Two-phase; b, three-phase. The third phase in b is a colloid or microparticle, shown here with contaminant molecules sorbed to it, thus making them mobile. Colloidal material is usually chemically similar to the stationary macroparticle phase.

bations or changes in flow velocity from pumping, and through *in situ* precipitation of supersaturated mineral phases. Colloids are removed from the aqueous phase by deposition on stationary macroparticles; it is the efficiency of deposition that regulates the facilitation of contaminant transport⁶.

Kersting and colleagues' analysis¹ of colloids isolated from pumped ground water clearly identified the source of the Pu as a single underground nuclear test site, the test well and detonation site being 1.3 km apart. The elimination of other detonation cavities and contamination as the source of Pu makes it evident that the Pu has been transported through the ground water by some process.

As delineated by Ryan and Elimelech⁵, however, three conditions must be met for defensible evidence that colloids have transported contaminants: first, colloids must be present; second, contaminants must associate with them; and third, the colloid-contaminant combination must move through the aquifer. The results of Kersting *et al.* qualitatively meet the first two conditions. But, as the authors point out, the possibility of sampling artefacts meant that they could not quantify some of the parameters needed for supporting the detailed assessment of colloid-facilitated Pu transport in their study. That is, the third condition has not been rigorously met. Nevertheless, their work clearly shows that a low-solubility contaminant travelled some way from the source, perhaps at or near the local groundwater flow velocity (1–80 m yr⁻¹).

Has the Gordian knot been cut? No, I do not think so, although a good slice has been taken out of it. The fundamental difficulty remains the gap between field observations and expectations based on bench-scale experiments and theory. For example, according to classical filtration theory, colloid transport should be relatively limited (tens of metres⁷ or less under typical subsurface conditions). Advances in incorporating⁸ macroparticle chemical heterogeneity and site blockage into models have helped to narrow the gap, but it

nonetheless remains wide.

The trouble with most field studies is that the systems are difficult to manipulate, and are often too large and heterogeneous to characterize accurately. There is a great need to develop meso-scale experimental systems (several metres to more than ten metres) for the careful evaluation of the effects of system heterogeneity on colloid transport and the testing of methods for scaling up from the bench to the field.

Given that the work by Kersting *et al.* shows that Pu may be transported considerable distances through groundwater systems, can one conclude that the colloidal transport of actinides provides a significant exposure pathway from nuclear testing and waste sites? Not really. The very properties of compounds that make them good candidates for colloid-facilitated transport — low solubility and high particle reactivity — limit the amount of contaminant that can be transported: colloids are both the means and the bottleneck. But we need to know more. T. H. Huxley⁹ had it that "It is the customary fate of new truths to begin as heresies and to end as superstitions [dogma]". In its evolution from heresy to dogma, colloid-facilitated contaminant transport has become a perceived truth, widely recognized, but rarely understood in detail. □

Bruce D. Honeyman is in the Environmental Science and Engineering Division, Colorado School of Mines, Golden, Colorado 80302, USA. e-mail: bhoneyman@mines.edu

1. Kersting, A. B. *et al.* *Nature* **397**, 56–59 (1999).
2. Buffle, J. & Leppard, G. G. *Environ. Sci. Technol.* **9**, 2169–2175 (1995).
3. McCarthy, J. F. & Zachara, J. M. *Environ. Sci. Technol.* **23**, 496–502 (1989).
4. Honeyman, B. D. & Santchi, P. S. *J. Mar. Res.* **47**, 951–992 (1989).
5. Ryan, J. N. & Elimelech, M. *Colloids Surf. A* **107**, 1–56 (1996).
6. Roy, S. B. & Dzombak, D. A. *Environ. Sci. Technol.* **31**, 656–664 (1997).
7. McDowell-Boyer, L. M., Hunt, J. R. & Sitar, N. *Wat. Resour. Res.* **22**, 1901–1921 (1986).
8. Song, L. & Elimelech, M. *J. Colloid Interf. Sci.* **167**, 301–313 (1994).
9. Huxley, T. H. in *Darwiniana: Essays by Thomas H. Huxley* (Appleton, New York, 1896).

Plant viruses

New tool for Swiss army knife

Ortrun Mittelsten Scheid

Like as a virus is demanding. Tiny in size, and with limited genetic material, a virus must successfully take over comparatively gigantic cells or multicellular organisms. Unlike symbiotic invaders it will face a hostile reception, and it is no surprise that viruses have evolved to carry simplified, multifunctional conquest weapons, resembling the sophisticated equipment in the pocket of 007's dinner suit. A good example of such a weapon is the 'helper component protease' (HC-Pro), which is found in the group of potyviruses that infect, and cause severe disease in, a spectrum of plant species. Three papers published in *Proceedings of the National Academy of Sciences*¹, *EMBO Journal*² and *Cell*³ now tell us something about why HC-Pro makes these viruses so successful.

As indicated by its name, HC-Pro was identified as the protease that cleaves a viral polyprotein into its functional subunits. But, like a Swiss army knife, it packs in more than just a cutting blade. It binds RNA and is involved in amplification of the viral RNA genome, spreading the virus from the site of infection and transmitting it from plant to plant via sap-sucking insects⁴. Although almost all plant viruses have proteins to do these jobs (albeit in different combinations), the new work¹⁻³ shows that HC-Pro has another function unknown in any other virus — it paralyzes an important plant defence mechanism that normally acts against viruses.

The mechanism in question is based on 'post-transcriptional gene silencing' (PTGS), whereby specific transcripts are degraded before they can be translated⁵. The process was initially observed in transgenic plants when, instead of being overexpressed, transgenes (as well as homologous endogenous genes) became silent. Cytoplasmic viral templates can be degraded by the same mechanism, especially if parts of the virus have been integrated into the plant's nucleus as transgenes. Indeed, silencing of viral gene expression by PTGS can explain 'recovery' of virus-infected plants which, despite initial viral replication and spread throughout the plant, develop new, virus-free shoots and leaves⁶.

Because RNA viruses replicate through double-stranded RNA intermediates, and silenced transgenes are suspected to produce aberrant antisense RNA, double-stranded RNA could be the common signal for specific degradation, probably also related to the mechanism that operates in animal models⁷. PTGS is thought to have evolved either as an antiviral defence mechanism, or as a regulatory mechanism for endogenous plant genes that also affects viral genes⁸. In both cases, the

potential to act against many viral sequences, combined with the high specificity of the response, seems to make PTGS part of a plant's 'immune system' based on particular RNA sequences. Therefore, it is not surprising that, like their animal counterparts, some plant viruses have developed strategies to attack this defence. The HC-Pro protein of tobacco etch virus and protein 2b of cucumber mosaic virus are part of this attack. Anandalakshmi *et al.*¹, Brigneti *et al.*² and Kasschau and Carrington³ now show that when these proteins are expressed in tobacco (*Nicotiana tabacum*) and *N. benthamiana*, the PTGS response is suppressed — existing silencing and virus resistance are removed, and silencing cannot be established during viral infection.

These results also bear on the concept of synergistic viral infection. When plants are infected with two different species of virus, the result is often more severe symptoms than inoculation with either virus alone. Potyviruses are successful partners in this double-attack strategy, allowing strong replication and spread of the other partner. HC-Pro is essential for double infection⁹, and it is probably its ability to suppress PTGS that limits the general plant defence response, allowing other viruses that do not encode HC-Pro to be amplified. Viral modulation of the plants' gene-silencing system might also cause unexpected changes in gene expression if transgenic plants cannot be saved from uncontrolled infections.

We do not yet know how, at the molecular level, HC-Pro is involved in viral amplification, spread and transmission. Is it really a

multifunctional tool, or are all of the functions that support infection simply consequences of suppressing PTGS? Whatever the answer, HC-Pro is a promising bait to fish for interacting plant components and to learn more about the molecular basis of PTGS. The newly discovered function of this viral Swiss army knife will make it a tool for plant geneticists, allowing gene silencing in transgenic plants to be increased or avoided. But because exaggerated euphoria about gene-transfer techniques has often been followed by unwanted gene silencing, we will probably also discover the limits to which we can manipulate gene silencing. Components such as HC-Pro are not the only factors involved, and they cannot provide a 100% effective control. Armed with the new information, it is amusing to note that the connection between gene silencing and viral resistance was first described for tobacco etch virus⁶, despite the fact that it carries HC-Pro and was the source of this protein for the new studies describing its PTGS-suppressing effect¹⁻³. In any case, it will be exciting to learn how plants combat this viral super-weapon from their side. □

Ortrun Mittelsten Scheid is at the Friedrich Miescher Institute, PO Box 2543, CH-4002 Basel, Switzerland.

e-mail: ortrun@fmi.ch

1. Anandalakshmi, R. *et al.* *Proc. Natl. Acad. Sci. USA* **95**, 13079–13084 (1998).
2. Brigneti, G. *et al.* *EMBO J.* **16**, 6739–6746 (1998).
3. Kasschau, K. D. & Carrington, J. C. *Cell* **95**, 461–470 (1998).
4. Maia, I. G., Haenni, A. L. & Bernardi, F. J. *Gen. Virol.* **77**, 1335–1341 (1996).
5. Depicker, A. & Van Montagu, M. *Curr. Opin. Cell Biol.* **9**, 373–382 (1997).
6. Lindbo, J. A., Silva-Rosales, L., Proebsting, W. M. & Dougherty, W. G. *Plant Cell* **5**, 1749–1759 (1993).
7. Fire, A. *et al.* *Nature* **391**, 806–811 (1998).
8. Covey, S. N., Al-Kaff, N. S., Lángara, A. & Turner, D. S. *Nature* **385**, 781–782 (1997).
9. Shi, X. M., Miller, H., Verchot, J., Carrington, J. C. & Vance, V. B. *Virology* **231**, 35–42 (1997).

Coral reefs

Recruitment in space and time

Peter F. Sale

Six years ago, in an address to the Ecological Society of America, Simon Levin wrote¹ that "the problem of pattern and scale is the central problem in ecology, unifying population biology and ecosystem science, and marrying basic and applied ecology". Interest in scale effects has since grown, and patterns or processes have been examined contemporaneously over several scales of space and time. On page 59 of this issue, Terry Hughes and colleagues² report an important field study on patterns of recruitment and adult abundance of corals on the Great Barrier Reef.

Corals, like most reef species, have a life cycle that combines a brief, dispersive, larval phase and a sedentary (in this case immobile)

adult phase. At the end of larval life, the animal must locate a suitable habitat in which to settle, metamorphose and begin its sedentary life. Ecologists term this arrival of new juveniles recruitment. For corals, suitable habitats occur in patches of reef separated by metres to thousands of kilometres of unsuitable, non-reef habitat. Transport of larval stages connects the many local populations of each species, and this interconnection drives local demography. Once marine ecologists woke up to the importance of recruitment, in the late 1970s, they found that its extent varied in both space and time^{3,4}.

Hughes and colleagues have now explored variability in recruitment of corals on four spatial scales and over two successive years



Figure 1 Corals can be divided into spawners, here seen releasing gametes during the mass spawning event, and brooders. Most of the species in the Great Barrier Reef are spawners, releasing packets of sperm or eggs for external fertilization. The resulting larvae may be pelagic for a week or more. The eggs of brooder species, however, are retained after fertilization until they are released at hatch. Hughes *et al.*² took advantage of the mass spawning to study patterns of larval recruitment and adult abundances, and conclude that recruitment rates vary substantially in location and time.

(1995 and 1996), and they relate the rates of recruitment at each spatial scale to the abundance of established adults. Scales ranged from that of reef sectors 250–500 km apart to that between replicates only metres apart. Unglazed ceramic tiles, individually bolted to the reef platform, provided standard surfaces on which recruiting corals settled. Nearby line transects provided data on adult corals. The authors were aided by the mass spawning of corals (Fig. 1). Most species of coral on the Great Barrier Reef spawn together over just one or two nights, soon after the full Moon in November each year⁵. This meant that tiles deployed at each site for a brief (eight-week) period sampled the total annual recruitment to that site. Despite this cnidarian cooperativeness, Hughes and colleagues had a logistically complex task, deploying and retrieving tiles at 72 sites arrayed over 2,000 km of the Great Barrier Reef and, in the first year, sampling the abundances of established corals at each site.

With this study, Hughes *et al.* illustrate the pattern of variation in recruitment, the spatial scales at which variation is strongest, and the extent to which recruitment mirrors patterns of adult abundance. The picture, although clear, is far from simple. It is also the first such picture for corals because, although coral recruitment has been studied since the mid-1980s, all previous work has been done at single, or localized, sites.

There are three related and important take-home messages. The first is that recruitment rates vary substantially among locations on the Great Barrier Reef. This variation is greatest at scales of 250–500 km (sectors

and 0.5–3 km (sites within reefs); it is much lower at the intermediate (between-reef) scale of 10–15 km. Recruitment rates also differ between spawner and brooder corals, and between years. The second message is that variation in recruitment does not match variation in the abundance of adults, indicating that other demographic processes also probably vary between locations. The third is that the other two messages assume that 1995 and 1996 were 'typical' years.

The third message needs to be made pointedly to those who fund ecological science. Hughes and colleagues' study is 'big' science, yet it has yielded only one small snapshot of the coral reef — this work must be repeated at other times and places. For those interested in the demography of corals, the other two messages provide a glimpse of the tangible realism that will be essential if we are to develop useful models of the dynamics of these complex open populations. These two messages are also very important for those who manage coral reefs.

Coral reefs rival rainforests in species diversity (supporting 1–9 million species), and greatly exceed them in the number of phyla (32 of 33) present⁶. As well as being important biodiversity stores, coral reefs are economically valuable to the (chiefly developing) countries that are privileged to have them. Reefs occupy less than 0.2% of the world's oceans, yet they provide 25% of the fishery catch in developing countries, plus other valuable exported products. Thriving reef environments also generate a substantial and growing revenue through tourism, and protect coastal communities from storms and marine erosion. These considerable economic and ethical values all require that the reef be managed sustainably, yet there is

growing awareness — and some alarm — that, worldwide, coral reefs are being seriously degraded by human activities^{7,8}. These include direct effects such as overfishing, and the less direct effects of pollution or, perhaps, global warming.

As a science reef management is in its infancy. We assess the condition of the reef by tedious surveys of coral cover, and similar snapshot approaches to the present state of the system. But by demonstrating that recruitment and adult population patterns are mismatched, Hughes and colleagues now show that current state need not predict current dynamics. Because a coral reef is a living system, we'd be wiser to base management on dynamics rather than state, because dynamics determine tomorrow's state.

Yes, unfortunately, the message to managers is this — your task, already difficult, just got tougher. The science that we must build to guide reef management will have many dimensions, and Hughes and colleagues have shown that the ecological dimension will not be trivial. □

Peter F. Sale is in the Biology Department and the Great Lakes Institute for Environmental Research, University of Windsor, Windsor, Ontario, Canada N9B 3P4.

e-mail: sale@uwindsor.ca

1. Levin, S. A. *Ecology* **73**, 1943–1967 (1992).
2. Hughes, T. P. *et al.* *Nature* **397**, 59–63 (1999).
3. Doherty, P. J. & Williams, D. McB. *Oceanogr. Mar. Biol. Annu. Rev.* **26**, 487–551 (1988).
4. Caley, M. J. *et al.* *Annu. Rev. Ecol. Syst.* **27**, 477–500 (1996).
5. Harrison, P. L. *et al.* *Science* **223**, 1186–1189 (1984).
6. Reaka-Kudla, M. L. in *Biodiversity II: Understanding and Protecting our Natural Resources* (eds Reaka-Kudla, M. L., Wilson, D. E. & Wilson, E. O.) 83–108 (Joseph Henry/National Academy Press, Washington DC, 1996).
7. Jackson, J. B. C. *Coral Reefs* **16**, S23–S32 (1997).
8. Dight, I. J. & Scherl, L. M. *Coral Reefs* **16**, S139–S147 (1997).

Photosynthesis

Making light of a dark situation

Robert Willows

When the seedlings of flowering plants are grown in the dark (etiolated seedlings), they cannot make chlorophyll because the enzyme that reduces protochlorophyllide to chlorophyllide — protochlorophyllide oxidoreductase (POR) — requires light. As well as collecting light for photosynthesis, chlorophylls protect the plant by dissipating excess light energy. So, when etiolated seedlings emerge from the soil into the sunlight they lack these photoprotective mechanisms. Moreover, if a seedling emerges into low light it may not be able to synthesize chlorophyll rapidly enough to take full advantage of this photosynthetic opportunity. On page 80 of this issue, Reinbothe *et al.*¹ show that POR, which presents the plant with these dilemmas, also provides it with the solution.

The pale green colour of etiolated seedlings is due to the accumulation of protochlorophyllide complexed with a reducing agent (NADPH) and POR in the developing chloroplast (etioplast; Fig. 1, overleaf). There are two differentially regulated PORs in the etioplast^{2,3}: PORA is present at a high level and is negatively regulated by light; PORB is found at lower, constitutive levels. Reinbothe *et al.*¹ have now found that, in barley, PORA and PORB form a stable complex in a ratio of 5:1, and they call this a light-harvesting POR–protochlorophyllide complex (LHPP). They also show that the PORA and PORB in these complexes are specific only for protochlorophyllide *b* and protochlorophyllide *a*, respectively.

Reinbothe and colleagues reconstituted LHPP *in vitro* and found that it had different

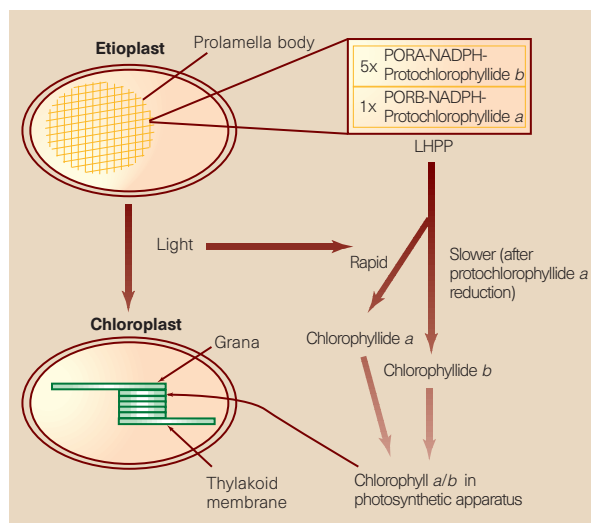


Figure 1 Conversion of an etioplast to a chloroplast. The interaction of a protochlorophyllide oxidoreductase (POR)–NADPH–protochlorophyllide complex with membranes results in the formation of prolamella bodies in the developing chloroplast (etioplast). Reinbothe *et al.*¹ have reconstituted this complex *in vitro*, and call it the light-harvesting POR–protochlorophyllide complex (LHPP).

spectral properties from the complex *in vivo*. However, by adding galactolipids and sulpholipids from etioplasts to the mixture, they were able to obtain LHPP with spectral characteristics similar to those of etioplast membrane preparations. They found that protochlorophyllide *b* in the LHPP initially has a light-harvesting function. It then transfers light energy to protochlorophyllide *a* for photoreduction to chlorophyllide *a* by PORB. By this transfer, protochlorophyllide *a* is rapidly photoreduced — even under low light intensities — so that the seedling can assemble its photosynthetic apparatus quickly.

Although protochlorophyllide *a* is reduced quickly, the protochlorophyllide *b* is not immediately photoreduced at the light-harvesting stage. Any excess light energy that is not transferred to protochlorophyllide *a* is then emitted as fluorescence. Only when the complex between protochlorophyllide *b* and PORA is separated from the LHPP can this protochlorophyllide be efficiently photoreduced; this means that, under a high light intensity, protochlorophyllide *b* has a photoprotective function. This ties in with the fact that there is five times as much protochlorophyllide *b* as *a* in the LHPP. Once enough chlorophyll has been made and assembled into a functional photosynthetic apparatus, PORA is no longer needed, explaining why this enzyme is negatively regulated by light.

Protochlorophyllide *b* has been identified in many fully green plants⁴ but, despite numerous spectroscopic studies⁵, it has never before been reported in etiolated tissue. So, the implication of Reinbothe and colleagues' study — that most of the protochlorophyllide in barley etioplasts is protochlorophyllide *b* — is not yet supported experimentally. However, some characteristics of chlorophyll-*b*-deficient mutants (which, presumably, cannot make protochlorophyllide *b*) may be explained by the requirement for protochlorophyllide *b* in the LHPP. These mutants almost universally cannot green rapidly when exposed to light⁶, and they do not develop well when greened

under high light conditions. These are the phenotypes that would be expected if PORA could not carry out its function in the LHPP.

Expression of PORA can be suppressed in the dicotyledon *Arabidopsis thaliana* if the plant is grown under continuous far-red light. These plants do not make prolamella bodies (etioplast precursors of the grana) and they cannot green rapidly⁷. However, when either PORA or PORB is overexpressed in these *Arabidopsis* seedlings⁸ (or in *Arabidopsis* mutants that lack prolamella bodies⁹), the prolamella bodies form and the seedlings green normally. These experiments^{8,9} indicate that the quantity of POR is important for formation of the prolamella body and greening in *Arabidopsis*, and they do not exclude the possibility that an LHPP containing a PORA–protochlorophyllide-*b* complex occurs in this plant. However, they do contradict the suggestion by Reinbothe *et al.*¹ that formation of the prolamella body requires both PORA and PORB in an LHPP.

The specificity of protochlorophyllide *b* for the highly expressed PORA in etiolated seedlings is important for formation of the LHPP, with its interesting photochemical properties. The next challenge will be to see whether the PORs of other plant species have similar substrate specificities, or whether the LHPP is unique to barley. □

Robert Willows is in the School of Biological Sciences, Macquarie University, Macquarie Drive, North Ryde, 2109 Sydney, Australia.
e-mail: rwillows@rna.bio.mq.edu.au

1. Reinbothe, C., Lebedev, N. & Reinbothe, S. *Nature* **397**, 80–84 (1999).
2. Armstrong, G. A., Runge, S., Frick, G., Sperling, U. & Apel, K. *Plant Physiol.* **108**, 1505–1517 (1995).
3. Holtorf, H., Reinbothe, S., Reinbothe, C., Berezina, B. & Apel, K. *Proc. Natl. Acad. Sci. USA* **92**, 3254–3258 (1995).
4. Shedbalkar, V. P., Ioannides, I. M. & Rebeiz, C. A. *J. Biol. Chem.* **266**, 17151–17157 (1991).
5. Adamson, H. Y., Hiller, R. G. & Walsley, J. *J. Photochem. Photobiol. B* **41**, 201–221 (1997).
6. Falbel, T. G. & Staehelin, L. *Physiol. Planta* **97**, 311–320 (1996).
7. Runge, S., Sperling, U., Frick, G., Apel, K. & Armstrong, G. A. *Plant J.* **9**, 513–523 (1996).
8. Sperling, U., van Cleve, B., Frick, G., Apel, K. & Armstrong, G. A. *Plant J.* **12**, 649–658 (1997).
9. Sperling, U. *et al.* *Plant Cell* **10**, 283–296 (1998).

Daedalus

Highly pointed remarks

The new global telephone systems, with their multitude of orbiting satellites, worry radioastronomers. The Iridium system now entering service uses 66 satellites; at least one, and usually several, are above the horizon at any time anywhere on Earth. The radio glare of their transmitters blots out the feeble radiation from cosmic sources. Daedalus now points out that radioastronomy itself has pioneered a solution to the problem.

The satellites are all in communication with one another; but only one (the nearest) communicates with any given client station on the ground. But suppose all the satellites in sight of that station joined in the communication. They would collectively form a 'very long baseline interferometer' aerial in the sky, spread over millions of square kilometres of orbital space. With the right phasings between its elements, such an aerial is wonderfully directional. It could transmit to, and receive from, a mere few square metres around a specific ground station. Receivers elsewhere, such as radio telescopes, would detect nothing.

Not only would this system eliminate interference, it would drastically reduce the power needed for its transmitters. For instead of broadcasting wastefully to whole regions, the satellite array could target each ground station exactly. Its sensitivity as a receiving array would be equally enhanced. Properly phased, the collective satellite aerial would have a very high 'gain' for signals from that station. So the satellites could be put into higher orbits, their horizons would broaden, and fewer of them would be needed to cover the globe. The only snag is that their positions are constantly changing. And to compute the correct phasings needed to target each ground station, and keep them properly updated, those positions would have to be accurately known all the time.

No problem, says Daedalus. The most accurately located satellites of them all are, of course, those of the Global Positioning System. Several of those must also be in the sky over any point at any time. So combine the two systems in one set of satellites. Fewer satellites would crowd the heavens, transmitter power and radio spectrum space would be saved, and radioastronomers everywhere could relax. They would merely have to avoid using the satellite telephone while at the telescope. Otherwise their cosmic gaze would be disrupted by a sudden accurately targeted blast of chatter from the sky.

David Jones

Timing the end of nocturnal sleep

Some people can quite accurately time the end of their night's sleep at will, without using an alarm clock¹, demonstrating that it is possible to voluntarily control a state of consciousness that is characterized by a loss of volition and attentional guidance². Here we show that the expectation that sleep will come to an end at a certain time induces a marked increase in the concentration of the hormone adrenocorticotropin in the blood one hour before waking. The regulation of adrenocorticotropin release during nocturnal sleep is therefore not confined to daily rhythms; it also reflects a preparatory process in anticipation of the end of sleep.

The regulation of sleep termination has been thought to be embedded in a daily (circadian) rhythm controlling in parallel the release of pituitary and adrenal hormones^{3,4}. Normally, the release of adrenocorticotropin and cortisol increases during late stages of sleeping, reaching a daily maximum at the time of spontaneous waking⁵. Adrenocorticotropin and cortisol are also released from the pituitary–adrenal system in a major adaptive response to stress, and are secreted in anticipation of stressful events^{6,7}. We investigated whether the increase in the secretion of pituitary–adrenal hormones during the late stages of sleeping in part reflects anticipation of the 'stress' of the waking phase.

Fifteen healthy volunteers (mean age, 25.2 years) with regular sleep–wake rhythms were studied during three nights. We made polysomnographical recordings (electroencephalogram, electrooculogram and electromyogram) throughout the night, and took blood samples every 15 minutes to determine plasma concentrations of adrenocorticotropin and cortisol.

Lights were turned off at midnight, after subjects had been told they would be woken at either 6:00 ('short sleep', on one night) or 9:00 ('long sleep', on the other two nights). On one of the long-sleep nights they were woken at 9:00 as they expected, but on the other night they were instead woken at 6:00 ('surprise') under the pretence of a technical problem. After being woken, subjects stayed in bed for another three hours. We interviewed the volunteers at the end of the experiments, and found that all but one of the subjects (we have not used data from this subject) had expected to be woken up at the specified time. The order of the three experimental nights was balanced across subjects, with five subjects starting with short sleep, five with long sleep, and five with the surprise condition.

Plasma concentrations of adrenocorticotropin and cortisol increased during sleep, and the rate of increase did not differ among experimental conditions until about 4:30 (Fig. 1). However, when anticipating being woken up at 6:00, subjects showed a distinct increase in adrenocorticotropin levels within the last hour before waking (resulting in an average adrenocorticotropin level of 37.3 ± 3.6 pg ml⁻¹, mean $\pm$ s.e.m.), compared with the same subjects under 'surprise' conditions (25.5 ± 2.7 pg ml⁻¹, $P < 0.005$) and 'long sleep' conditions (26.5 ± 4.1 pg ml⁻¹, $P < 0.05$).

The increase in adrenocorticotropin before being woken in the 'short sleep' condition was not accompanied by a significant increase in cortisol concentration, which remained unchanged before 6:00 by the anticipation of waking. This finding cannot be explained only by cortisol release being unresponsive to adrenocorticotropin stimulation: an adrenocorticotropin-independent

mechanism of adrenal regulation may also be involved⁸.

In the morning, arousal from sleep induced a temporary increase in the concentrations of adrenocorticotropin and cortisol, which peaked about 30 minutes later, probably as an adaptive response to the stress of waking⁵. Compared with the level just before being woken, the increase in adrenocorticotropin levels after waking in the 'surprise' condition (an increase of 22.1 ± 3.9 pg ml⁻¹) was larger than the increase after both 'short sleep' (10.6 ± 3.2 pg ml⁻¹, $P < 0.01$) and 'long sleep' (12.2 ± 3.3 pg ml⁻¹, $P < 0.05$). A greater increase also occurred in cortisol concentrations after 'surprise' awakening (73.4 ± 6.6 ng ml⁻¹) than after awakening from 'short sleep' (40.4 ± 9.9 ng ml⁻¹, $P < 0.025$) or 'long sleep' (35.4 ± 9.8 ng ml⁻¹, $P < 0.005$).

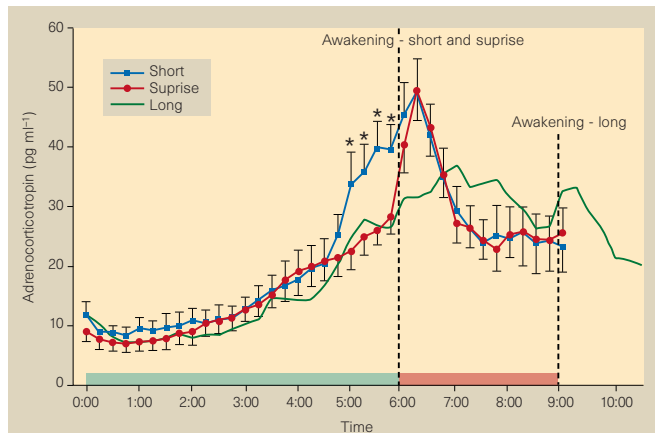
Our polysomnographical recordings did not show any differences in the nature of the sleep between 0:00 and 6:00 that depended on the anticipation of sleep termination, in agreement with previous findings⁹. In particular, 'short sleep' was similar to sleep under 'surprise' conditions. The proportions of time spent in the different sleep stages during the 'short sleep' and 'surprise' conditions were, respectively: sleep stage 1, $8.56 \pm 0.27\%$ and $7.10 \pm 1.72\%$; sleep stage 2, $35.40 \pm 1.38\%$ and $35.72 \pm 2.20\%$; slow-wave sleep, $37.05 \pm 3.68\%$ and $37.69 \pm 3.59\%$; rapid eye movement (REM) sleep, $17.36 \pm 1.14\%$ and $18.88 \pm 1.38\%$; and wake time, $1.62 \pm 0.82\%$ and $0.59 \pm 0.04\%$. Moreover, the temporal distribution of sleep stages (including transient arousals) did not differ between the 'short sleep' and 'surprise' conditions. Final waking occurred from REM sleep in five subjects of each condition.

The increase in adrenocorticotropin release before the expected time of waking indicates that anticipation, which is generally considered to be a unique characteristic of the regulation of conscious action, pervades sleep. The anticipatory adrenocorticotropin increase may also facilitate spontaneous waking¹⁰. The regulation of adrenocorticotropin release points to a mechanism that quickly adjusts endocrine activity to sharp changes in the duration of sleep.

Jan Born*, Kirsten Hansen*, Lisa Marshall*, Matthias Mölle*, Horst L. Fehm†

Departments of *Clinical Neuroendocrinology and †Internal Medicine, University of Lübeck, Ratzeburger Allee 160, Haus 23a, 23538 Lübeck, Germany
e-mail: born@kfg.mu-luebeck.de

Figure 1 Mean plasma concentrations of adrenocorticotropin for subjects under 'short sleep', 'surprise' and 'long sleep' conditions (see text). Standard error of the mean is shown for 'short sleep' and 'surprise' conditions. Broken vertical lines indicate waking. Asterisks indicate statistically significant differences between 'short sleep' and 'surprise' conditions. The increase in adrenocorticotropin before the end of 'short sleep' was also significant compared with levels during 'long sleep' in the hour before 6:00. The low adrenocorticotropin before 'surprise' waking, compared with levels before waking from 'short sleep', was fully compensated by increased hormone release after waking, so adrenocorticotropin levels in these conditions were identical 30 minutes after waking.



The increase in adrenocorticotropin before the end of 'short sleep' was also significant compared with levels during 'long sleep' in the hour before 6:00. The low adrenocorticotropin before 'surprise' waking, compared with levels before waking from 'short sleep', was fully compensated by increased hormone release after waking, so adrenocorticotropin levels in these conditions were identical 30 minutes after waking.

1. Moorcroft, W. H., Hennager Kayser, K. & Griggs, A. J. *Sleep* **20**, 40–45 (1997).
2. Hobson, J. A. & Stickgold, R. in *The Cognitive Neurosciences* (ed. Gazzaniga, M.) 1373–1389 (MIT Press, Cambridge, MA, 1995).

3. Van Cauter, E. & Turek, F. W. in *Endocrinology* Vol. 3 (ed. DeGroot, L. J.) 2487–2548 (Saunders, Philadelphia, 1995).
4. Czeisler, C. A., Weitzman, E. D., Moore-Ede, M. C., Zimmerman, J. C. & Knauer, R. *Science* **210**, 1264–1267 (1980).
5. Späth-Schwalbe, E., Schöller, T., Kern, W., Fehm, H. L. & Born, J. *J. Clin. Endocrinol. Metab.* **75**, 1431–1435 (1992).
6. Mason, J. W. *et al. Psychosom. Med.* **35**, 406–414 (1973).
7. Deinzer, R., Kirschbaum, C., Gesele, C. & Hellhammer, H. *Physiol. Behav.* **61**, 507–511 (1997).
8. Ehrhart-Bornstein, M., Hinson, J. P., Bornstein, S. R., Scherbaum, W. A. & Vinson, G. P. *Endocr. Rev.* **19**, 101–143 (1998).
9. Lavie, P., Oksenberg, A. & Zomer, J. *Percept. Motor Skills* **49**, 447–450 (1979).
10. Follenius, M., Brandenberger, G. & Muzet, A. *Horm. Metab. Res.* **17**, 602–606 (1985).

Air traffic may increase cirrus cloudiness

High-level cirrus clouds can evolve^{1,2} from the condensation trails of aircraft, which form as the mixture of warm, humid exhaust gases and colder, drier air exceeds water saturation³. In addition, the particles in exhaust plumes from aircraft may allow ice nucleation at lower supersaturations than those required under natural conditions⁴. This mechanism is sensitive to environmental conditions, but may occur downstream of the exhaust aerosol source regions. Here I show that cirrus clouds increased in occurrence and coverage in the main air-traffic flight corridors between 1982 and 1991.

I used synoptic cloud reports from land and ship stations for 1982 to 1991 (ref. 5) to establish the relationship between variations in the occurrence of cirrus cloud and the spatial distribution of aviation fuel consumption^{6,7}. During this period there was a large increase in total fuel consumption by air traffic, with an average annual growth rate of 3.2% (ref. 8). I computed the cirrus occurrence frequency at the $3^\circ \times 3^\circ$ resolution from cloud reports (see Supplementary Information). The change in cirrus occurrence frequency is then computed as the difference in frequency between 1987–1991 and 1982–1986, whereas its trend is estimated as the absolute percentage increase along a best-fit line over the 10 years of data.

The average change in cirrus occurrence over land and ocean is plotted against fuel consumption in Fig. 1a, b. Global annual cirrus occurrence increases on average by 1.1% and 3.5% for land and ocean, respectively, with regional average increases of 2.9% to 4.6% over the principal flight corridors (defined as the twenty $3^\circ \times 3^\circ$ grid-boxes with most air traffic). The largest changes correspond to positive trends in cirrus occurrence of 8.5% and 7.0% (Table 1). Over North America, the trend in cirrus occurrence varies between 2.4% and 9.9% per decade, depending on the season. It is largest in December–May, when there is the highest frequency of persistent condensation trails⁹. Between 39° and 42° N, close to

Table 1 Trends in cirrus cloud occurrence, cirrus amount when present and cirrus amount

	Occurrence				AWP		Amount
	DJF	MAM	JJA	SON	Annual	Annual	
Land	2.5/6.4	1.8/9.7	2.4/6.8	0.6/2.2	1.7/8.5	–1.9/–1.9	0.0/2.9
Eurasia	1.2/2.6	–0.8/1.1	1.7/4.4	–0.8/–1.4	0.6/1.6		
N. America	7.0/7.8	9.9/16.5	3.2/9.5	2.4/7.3	5.6/13.3		
Ocean	6.0/8.3	5.5/6.3	6.4/8.7	7.7/6.0	6.2/7.1	0.4/–4.2	2.3/1.0
N. Atlantic	4.5/8.4	5.1/6.3	3.3/8.7	3.9/6.0	4.1/7.1		
N. Pacific	8.7/5.8	8.4/1.9	5.8/2.5	10.4/4.8	8.4/4.1		

In each column, the first figure reports the trend over the whole region, and the second figure shows the trend over the 20 grid-boxes with most air traffic in 1992. For North America, the second figure is for the 10 grid-boxes with most air traffic. Months (December to November) are denoted by their initial letter; AWP, amount when present. Trends are given as per cent per decade.

the Great Lakes, the trend in annual mean cirrus occurrence is 13.3%, more than twice that for the rest of North America. Over ocean, the trend is also very large in the North Atlantic flight corridor, hence the positive slope in Fig. 1c. Changes in cirrus occurrence over the North Atlantic flight corridor are significantly larger than over the rest of the North Atlantic (Student's *t*-test, $P=0.03$). Table 1 indicates that, with only one exception, the trend in cirrus increase is larger averaged over the 20 grid-boxes in each region with most air traffic for Eurasia, North America and the North Atlantic Ocean in all seasons.

In the North Pacific Ocean there is the largest increase in cirrus occurrence between 40° and 60° N. There is little air

traffic here, but the region lies downwind of the high-traffic area along the east coast of Asia. It is plausible that cirrus occurrence in the North Pacific is favoured by the presence of soot aerosols emitted by aircraft and transported eastwards to areas where conditions for cirrus formation are often met; a large amount of natural cirrus cloud is found over this region. Along the east coast of Asia there has been a decrease in cirrus occurrence, which might be explained by a decrease in the mean flight level (computed from refs 6,7), which lowers the potential for condensation-trail formation.

A search for other possible causes of increased cirrus, such as the effects of the El Chichón and Mount Pinatubo volcanic aerosols¹⁰, or long-term changes in relative

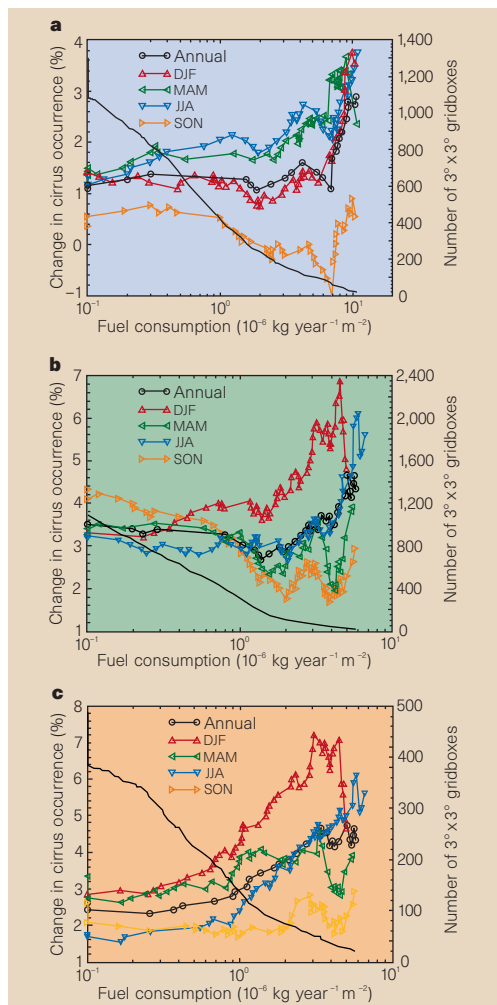


Figure 1 Difference in cirrus occurrence frequency between 1987–1991 and 1982–1986. **a**, Land; **b**, ocean; **c**, North Atlantic Ocean. Results are averaged over grid-boxes with the 1992 fuel consumption at altitudes more than 8 km greater than the values displayed on the x-axis, and show annual means (circles) and seasonal means (triangles). The frequency of cirrus occurrence is defined as the ratio of the number of high-cloud observations to the number of potential high-level cloud observations (situations where low- and middle-level clouds do not completely obscure the higher level of the atmosphere). Only observations meeting the illuminance criterion defined by Hahn *et al.*¹⁵ are included, introducing a bias towards daytime. The bold line represents the number of $3^\circ \times 3^\circ$ grid-boxes from which the annual mean is calculated. The curves are truncated when the number of grid-boxes used to calculate the mean falls below 20.

humidity and climate variations related to the North Atlantic oscillation^{11,12}, showed that none of them, taken individually, is a good explanation for the observed positive trend in cirrus occurrence and its regional distribution.

The trends in cirrus 'amount when present' over the previously defined regions of high fuel consumption are -1.9% and -4.2% for land and ocean, respectively (Table 1). The combination of a large positive trend in cirrus occurrence associated with a negative trend in the cirrus amount when present is consistent with an increasing amount of condensation-trail cirrus covering a small portion of an otherwise clear sky at the cirrus cloud level. This results in a trend in cirrus coverage (cirrus occurrence $\times$ cirrus amount when present) of 0.0% averaged over land, rising to 2.9% over the continental regions with most air traffic. Over ocean, these values are 2.3% and 1.0% , respectively, because of the aforementioned contribution of the North Pacific Ocean to the global increase in cirrus occurrence. Both radiative transfer calculations and Earth Radiation Budget Experiment measurements of the net change in radiation associated with high, thin clouds¹³ suggest a 0.2 W m^{-2} cloud radiative forcing per 1% of condensation trail or high-level cloudiness. I therefore estimate that the trend in cirrus amount over the continental regions of high air traffic (2.9% , as discussed above) would correspond locally to an additional cloud radiative forcing of about 0.7 W m^{-2} between 1982 and 1991. A larger forcing would be expected if calculated from the beginning of commercial jet traffic until the present time. However, we cannot rule out the possibility that the observed change in cirrus amount is not due solely to aviation, or that other climate processes have not masked an even larger aviation impact on cirrus cloudiness. For instance, a previous analysis¹⁴ showed that other regions experienced large positive trends in cirrus amount between 1952 and 1981.

Olivier Boucher

Laboratoire d'Optique Atmosphérique,
U.F.R. de Physique, Université de Lille-I,
59655 Villeneuve d'Ascq Cedex, France
e-mail: boucher@loa.univ-lille1.fr

- Seinfeld, J. H. *Nature* **391**, 837–838 (1998).
- Minnis, P. et al. *Geophys. Res. Lett.* **25**, 1157–1160 (1998).
- Schumann, U. *Meteorol. Z.* **5**, 4–23 (1996).
- Jensen, E. J. & Toon, O. B. *Geophys. Res. Lett.* **4**, 249–252 (1997).
- Hahn, C. J., Warren, S. G. & London, J. *Edited Synoptic Cloud Reports from Ships and Land Stations Over the Globe, 1982–1991*, NDP026B (Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Tennessee, 1996).
- Baughcum, S. L., Henderson, S. C., Tritz, T. G. & Pickett, D. C. *Scheduled Civil Aircraft Emission Inventories for 1992: Database Development and Analysis* (NASA CR-4700, 1996).
- Mortlock, A. M. & van Alstyne, R. *Military, Charter, Unreported Domestic Traffic and General Aviation: 1976, 1984, 1992, and 2015 Emission Scenarios* (NASA CR-207639, 1998).
- Energy Statistics and Balances of Non-OECD Countries, 1994–1995 124–125 (International Energy Agency, Organization for Economic Cooperation and Development, Paris, 1997).

- Minnis, P., Ayers, J. K. & Weaver, S. P. *Surface-Based Observations of Contrail Occurrence Frequency over the U. S., April 1983–April 1994* (NASA Reference Publication 1404, 1997).
- Jensen, E. J. & Toon, O. B. *Geophys. Res. Lett.* **19**, 1759–1762 (1992).
- Hurrell, J. W. *Science* **269**, 676–679 (1995).
- Möchel, H., Kapala, A. & Flohn, H. *Int. J. Climatol.* **18**, 1–22 (1998).
- Hartmann, D. L., Ockert-Bell, M. E. & Michelsen, M. L. *J. Clim.* **5**, 1281–1304 (1992).
- Warren, S. G., Hahn, C. J., London, J., Chervin, R. M. & Jenne, R. L. *Global Distribution of Total Cloud Cover and Cloud Type Amounts over the Ocean* (NCAR Technical Note TN-317 + STR, Boulder, Colorado, 1988).
- Hahn, C. J., Warren, S. G. & London, J. *J. Clim.* **8**, 1429–1446 (1995).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

High metabolic rates in running birds

The ability to increase metabolic rate during locomotion has been important in the structural evolution and evolutionary success of both birds and mammals. Greater endurance capabilities are conferred directly by greater maximal metabolic rates, which vary between species. These maximal rates are known for many mammals¹ but have not been determined for birds. We have measured oxygen consumption in a large flightless bird, the rhea, *Rhea americana*, while it was running on an inclined treadmill, and find an upper limit to aerobic metabolism that is 36 times greater than the minimum resting rate, a factorial increase exceeding that reported for nearly all mammals.

We trained two female rheas (average mass 21.8 kilograms) for two years to run on a treadmill while wearing clear plastic hoods. We then measured linear increases in rates of oxygen uptake while the birds ran up a 16° incline. The maximum rate of oxygen uptake was 2.85 millilitres per kilogram of body mass per second, reached at a running speed of 4.0 metres per second. There were no further increases in oxygen uptake when running speeds were increased to 4.7 metres per second, and plasma lactate concentrations indicated an increasing reliance on anaerobic glycolysis to provide metabolic energy at these higher speeds.

Rates of lactate accumulation (4.4 to 6.8 mM min^{-1}) and peak concentrations (24.8 mM) matched values reported for mammals running at speeds above their aerobic maximum. Remarkably, the aerobic scope

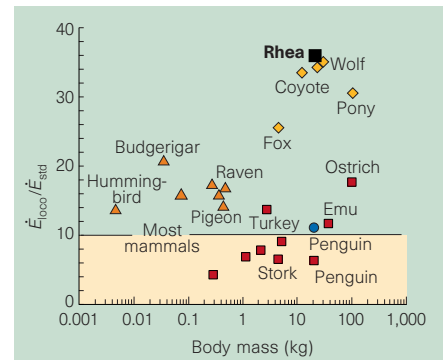


Figure 1 Factorial increases in metabolic rates above resting minimum for rheas and athletic mammals at the aerobic maximum and for other avian species at the highest rates available. Triangles, flying birds; squares, walking and running birds; circles, swimming birds; and diamonds, running mammals.

(the maximum aerobic metabolic rate divided by the minimum resting rate²) of these relatively inactive birds is more than 3.5 times those of typical mammals¹, and equals those of some of the most capable mammalian endurance runners (Fig. 1).

Although the rheas' factorial increase in anaerobic metabolism is 1.7 times greater than the highest previously reported avian values (Fig. 1), we believe that their aerobic scope is representative of birds in general. Previous avian values do not provide evidence of being upper limits; they are probably below species' maxima, but to unknown extents. Nonetheless, the minimum metabolic increase required for flight³ of 15 times the resting value exceeds the aerobic power of a typical mammal by a factor of 1.5, and most avian species can fly. Even birds with more limited endurance, such as the penguin and domestic fowl, can increase their aerobic metabolism by as much as a typical mammal^{4,5}. We suggest that, on average, the aerobic scope of birds exceeds those of mammals by a factor of two or more.

The structural basis of the rheas' aerobic power indicates that respiratory structure–function relationships, previously undetermined for birds, are quantitatively similar to those of mammals throughout most of the respiratory system, and are consistent with the concept of reasonable animal design⁶ (Table 1). Rheas, like aerobic mammals, achieve maximal high rates of oxygen uptake with existing cardiovascular and muscular structures, rather than by

Table 1 Mitochondrial volume densities and capillary densities of rhea muscles

Muscle	Bird A		Bird B	
	V_m (%)	N_c (mm^{-2})	V_m (%)	N_c (mm^{-2})
Iliotibialis lateral	10.1	803	6.5	787
Gastrocnemius	11.8	864	7.9	642
Pectoralis	4.2	411	4.2	469
Humerotriceps	4.3	619	4.9	508

Mean mitochondrial volume densities (V_m) and capillary densities (N_c) for leg muscles (iliotibialis and gastrocnemius) were 9.04% and 774 mm^{-2} , respectively; values for wing muscles (pectoralis and humerotriceps) were 4.37% and 502 mm^{-2} , respectively, for the rheas. Values for equivalent leg muscles of similarly sized mammals: for a 28-kg dog, $V_m = 9.1\%$ and $N_c = 884 \text{ mm}^{-2}$; for a 26-kg goat, $V_m = 3.4\%$ and $N_c = 719 \text{ mm}^{-2}$.

qualitative adjustment. Rhea heart mass (267 grams) conforms to that predicted for a mammal of the same body size and aerobic power. Volume densities of mitochondria in leg muscles are similar to those of mammals of the same size and aerobic capacity, whereas those in the relatively inactive flight muscles are only half as large and correspond to values reported for less aerobic mammals (Table 1).

Given that 30% of the rhea's body mass consists of leg musculature, rates of oxygen uptake by mitochondria per millilitre during locomotion at the aerobic maximum appear to fall within the range reported for mammals. Capillary densities in the rhea leg and flight muscles, like the mitochondrial volume densities, also parallel values reported for the muscles of athletic and less mobile mammals.

In contrast to the apparent conservation of structure-function relationships in most of the respiratory system, our results suggest that there are basic differences in the performance of the lungs of birds and mammals. We could not measure the lung volumes of the rheas directly, but in birds these are normally slightly more than half of those of mammals of the same mass^{7,8}. The rheas therefore achieved maximum oxygen flux rates, equal to those of the most aerobic mammals of their size, using lungs that are probably only half as large. This supports the general belief that avian lungs provide relatively more function per unit volume than mammalian lungs^{8,9}.

Although the aerobic limits of rheas and athletic mammals are similar, the metabolic power available in practice, and their functional needs, are not. Unlike dogs, horses and other athletic mammals that sustain high metabolic rates for hours during predation and migration, rheas do little or no sustained running and are poor at dissipating metabolic heat loads¹⁰. Rheas have apparently not been under strong selective pressures like those that promoted the aerobic power of extant running mammals. Large flightless birds lead fairly inactive lives, and may have lost the ability to fly primarily because of a lack of predation. Why rheas have an aerobic power that greatly exceeds their apparent functional needs remains a puzzle.

Matthew W. Bundle*, **Hans Hoppeler†**,
Ruth Vock†, **June M. Tester***,
Peter G. Weyand*

*Museum of Comparative Zoology,
Harvard University,
Bedford, Massachusetts 01730, USA
e-mail: pweyand@oeb.harvard.edu

†Department of Anatomy,
University of Bern,
Buhlstrasse 26, CH 3000 Bern 9, Switzerland

- Seeherman, H. J., Taylor, C. R., Maloiy, G. M. O. & Armstrong, R. B. *Respir. Physiol.* **44**, 11–23 (1981).
- Lasiewski, R. C. & Dawson, W. R. *Condor* **69**, 13–23 (1967).

- Ellington, C. P. *J. Exp. Biol.* **160**, 71–91 (1991).
- Brackenbury, J. H. & El-Sayed, M. S. *J. Exp. Biol.* **117**, 349–355 (1985).
- Kooyman, G. L. & Ponganis, P. J. *J. Exp. Biol.* **195**, 199–209 (1994).
- Weibel, E. R., Taylor, C. R. & Hoppeler, H. *Proc. Natl Acad. Sci. USA* **88**, 10357–10361, (1991).
- Maina, J. N. & King, S. A. *J. Anat.* **163**, 67–73 (1989).
- Dubach, M. *Respir. Physiol.* **46**, 43–60 (1981).
- Piiper, J. & Scheid, P. in *Comparative Physiology* (eds Bolis, L., Schmidt-Nielsen, K. & Maddrell, S. H. P.) 161–185 (North-Holland, 1973).
- Taylor, C. R., Dmi'el, R., Fedak, M. & Schmidt-Nielsen, K. *Am. J. Physiol.* **51**, 772–776 (1981).

The Thomas Jefferson paternity case

The DNA analysis of Y-chromosome haplotypes used by Foster *et al.*¹ to evaluate Thomas Jefferson's alleged paternity of Eston Hemings Jefferson, the last child of his slave Sally Hemings, is impressive. However, the authors did not consider all the data at hand in interpreting their results.

No mention was made of Thomas Jefferson's brother Randolph (1757–1815), or of his five sons^{2,3}. Sons of Sally Hemings conceived by Randolph (or by one of his sons) would produce a Y-chromosome analysis identical to that described by Foster *et al.* Further collaborative data (for example, the whereabouts of any of those who might have been involved at conception) are needed to confirm that Jefferson did indeed father his slave's last child, as claimed in the title. We know Thomas Jefferson was there, but how about Randolph Jefferson and his sons?

David M. Abbey

Health Science Center, University of Colorado,
1512 Teakwood Court, Fort Collins,
Colorado 80525, USA
e-mail: dabbey1000@aol.com

If the data of Foster *et al.* are accurate, then any male ancestor in Thomas Jefferson's line, white or black, could have fathered Eston Hemings. Plantations were inbred communities, and the mixing of racial types was probably common. As slave families were passed as property to the owner's offspring along with land and other property, it is possible that Thomas Jefferson's father, grandfather or paternal uncles fathered a male slave whose line later impregnated another slave, in this case Sally Hemings. Sally herself was a light mulatto, known even at that time to be Thomas Jefferson's wife's half sister^{3,4}.

Gary Davis

Evanston Hospital, 2650 Ridge Avenue,
Evanston, Illinois 60201, USA

Foster et al. reply — It is true that men of Randolph Jefferson's family could have fathered Sally Hemings' later children. Space constraints prevented us from expanding on alternative interpretations of

our DNA analysis, including the interesting one proposed by Davis. The title assigned to our study was misleading in that it represented only the simplest explanation of our molecular findings: namely, that Thomas Jefferson, rather than one of the Carr brothers, was likely to have been the father of Eston Hemings Jefferson.

It had been suggested to us earlier (by Herbert Barger, who also helped to recruit the descendants of Field Jefferson who participated in our study) that Isham Jefferson, son of Thomas Jefferson's brother Randolph, might have been the father of one or more of Sally Hemings' children. Barger's proposal was based on a statement⁵ that Isham was reared by Thomas Jefferson; he suggested that Isham could have been at Monticello or at Snowden (Snowden was across the James River from Scottsville, which is about 20 miles from Monticello) when Eston Hemings was conceived. But it is not known for certain that Isham was at Monticello at that time, whereas it is documented that Thomas Jefferson was. From the historical knowledge we have, we cannot conclude that Isham, or any other member of the Jefferson family, was as likely as Thomas Jefferson to have fathered Eston Hemings.

We know from the historical and the DNA data that Thomas Jefferson can neither be definitely excluded nor solely implicated in the paternity of illegitimate children with his slave Sally Hemings. When we embarked on this study, we knew that the results could not be conclusive, but we hoped to obtain some objective data that would tilt the weight of evidence in one direction or another. We think we have provided such data and that the modest, probabilistic interpretations we have made are tenable at present.

E. A. Foster*, **M. A. Jobling†**, **P. G. Taylor†**,
P. Donnelly‡, **P. de Knijff§**, **R. Mieremet§**,
T. Zerjal¶, **C. Tyler-Smith¶**

*6 Gildersleeve Wood, Charlottesville,
Virginia 22903, USA
e-mail: eafoster@aol.com

†Department of Genetics,
University of Leicester, Adrian Building,
University Road, Leicester LE1 7RH, UK

‡Department of Statistics,
University of Oxford, South Parks Road,
Oxford OX1 3TG, UK

§MGC Department of Human Genetics,
Leiden University, PO Box 9503,
2300 RA Leiden, The Netherlands

¶Department of Biochemistry,
University of Oxford, South Parks Road,
Oxford OX1 3QU, UK

- Foster, E. A. *et al. Nature* **396**, 27–28 (1998).
- Mayo, B. & Bear, J. A. Jr *Thomas Jefferson and his Unknown Brother* (Univ. Press of Virginia, Charlottesville, 1981).
- Brodie, F. M. *Thomas Jefferson: An Intimate History* (Norton, New York, 1974).
- Randall, W. S. *Thomas Jefferson: A Life* (Holt, New York, 1993).
- History of Todd County, Kentucky* (1884).

Starry, starry night

The Heavens on Fire: The Great Leonid Meteor Storms

by Mark Littmann

Cambridge University Press: 1998. 348 pp.
£25, \$39.95

Owen Gingerich

Early on the morning of 13 November 1833, residents of eastern North America who happened to be awake (or awakened!) were treated to an astonishing, unexpected and fearsome sight. The sky was filled with shooting stars more numerous than one could count, a veritable storm of meteors.

Historical researches in the days that followed revealed that the event was not as singular as had first been supposed. A similar display had been seen from Venezuela by Alexander von Humboldt 34 years earlier, on 12 November 1799. And another interesting fact emerged: based on a bright meteor observed from two locations, the Yale mathematics professor Denison Olmsted calculated that it had burned out 30 miles above the Earth.

The meteorologists were offended. Meteors were their turf, and they thought they knew that the atmosphere did not extend that high. The notion that the meteors might be periodic smacked of astronomy. How absurd!

By 1866 meteors were well established as cosmic phenomena, and astronomers keenly awaited another meteor storm. This time it was Europe's turn for a dazzling display, though by all accounts not one to match the 1833 spectacle. And before the year was out, the Italian astronomer Giovanni Schiaparelli matched the orbit of another meteor shower, the annual August shooting stars that seemed to radiate from the constellation Perseus, with that of the Comet 1862 III. He could not, however, identify the November meteors (by this time called the Leonids) with any known comet. The reason was that the comet, Tempel-Tuttle, 1866 I, had only just been found, and its orbit was not immediately known.

Needless to say, astronomers eagerly awaited a repeat performance in 1899. The press played the story with a flourish, and turned upon the scientists with scorn when the great show did not materialize. Jupiter, the real villain, had perturbed the comet and its stream of dusty debris. No substantial performance arrived with 1933, and then — surprise! — a grand storm in 1966 for those in the western United States stubborn enough to have stayed up into the wee hours of the morning of 17 November.

Now, to have made a book of all this requires a fertile and informed imagination, and this Mark Littmann, professor of science journalism at the University of Tennessee, has brought to his project in abundance.



High expectations: the *Denver Post* reports the failed appearance of the predicted 1899 meteor storm.

Littmann tells his story with real flair, and expansively enough to teach a great deal of meteor, comet and meteorite astronomy in the process. There is considerable technical detail, mostly in side-bars, so the book is altogether satisfying both as a quick, informative read and also as a reference source. In addition, it is filled with well-chosen illustrations.

For those who would like to observe some meteors, and even to photograph them, the instructions are here in Littmann's book. Showers come every year, but storms rarely. Still, the Leonids are not the only sequence going. I vividly remember catching the Giacobini-Zinner storm in 1946, counting 1,500 meteors in three hours despite bright moonlight — this is a conveniently placed shower that comes in the evening rather than in the morning skies (and Littmann explains why).

Two aspects of the book require special comment. First, in treating the historical material, Littmann frequently adds in italics our contemporary view of the matter. Professional historians of science tend to frown on this as whiggishness, but I found it entirely appropriate for a general audience. Second, the book is overfilled with grey side-bars enriching the historical or technical material. The book's designer may have thought they made quite beautiful patterns, but as a reader

I found them maddeningly distracting, especially when they broke up the narrative flow right in the middle of sentences.

Another anniversary of Comet Tempel-Tuttle is now upon us. After being lost for a century, and presumed to have disintegrated into the orbital debris responsible for the meteor shower, the comet was rediscovered in 1966. It again faded into invisibility on its long trip out past Uranus, but on this round it was caught nearly a year before its perihelion passage (closest approach to the Sun) by astronomers using the 10-metre Keck telescope atop Mauna Kea in Hawaii. Last November, some observers viewed a satisfying display, though hardly a meteor storm.

Perhaps 1999 will furnish a more memorable show. If so, it will be the last good opportunity in a century, because the placement of the stream will not be favourable again until 2098, and after that 2131 and 2164 are not as promising. "Those are the Leonids' last gasp," Littmann writes, for the orientation of the orbit will slowly drift away from the Earth. There were no great Leonid storms before the ninth century, and there will be no more beyond 2164. □

Owen Gingerich is at the Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, USA.

The wizard of Menlo Park

Edison: A Life of Invention

by Paul Israel

Wiley: 1998. 552 pp. \$30, £19.50

Willem Hackmann

The life of Thomas Alva Edison has all the ingredients of which Samuel Smiles, the well-known Victorian biographer of famous engineers, would have approved. Smiles's portrait of Edison would have been of a solitary inventor who, from humble beginnings, achieved greatness by dint of hard work and native intelligence; a man making the most of his opportunities in the best Victorian tradition. Indeed, so much has been written about Edison, the 'wizard of Menlo Park', that the real person has all but disappeared under a mountain of paper.

Paul Israel in this latest biography has done a remarkable job. Not only has he given us fresh insights into a complex personality, but he has set this against the backdrop of a dramatically changing American society driven on remorselessly by the second Industrial Revolution in which Edison was a pivotal player. What makes this book so authoritative is the unprecedented access the author has had to the more than five million pages (workshop diaries, notebooks and letters) that make up the Thomas A. Edison Papers at Rutgers University. The 18-year project to produce a multivolume documentary edition of these papers has been distilled into this detailed biography. Almost literally, no stone has been left unturned in the search for new and corroborating material. Thus, an archaeological dig at Edison's home turned up a wide variety of fragments of chemical glassware, supporting evidence for his fascination with chemistry that persisted throughout his career.

It was a chemical accident that may well have led to Edison's deafness. While working as a newsboy on the Grand Trunk Railway, he was allowed to carry out experiments in the baggage car of the train until a bottle of phosphorus broke and set the car on fire. The severe boxing his ears were given by the baggage master after the fire had been put out may well have started the hearing impairment from which he suffered increasingly in later life. At times, however, he found his selective hearing quite useful.

Israel shows convincingly that a key aspect of Edison's scientific practice was his ability as a practical bench chemist. Often, to resolve a problem, his technique was to 'ransack nature's warehouse'; examples filled his laboratories. When he had trouble finding a suitable filament for his incandescent lamp in the 1880s, he determined, after testing numerous natural substances, that maduke



A man of many hats: by the end of his industrious career, Edison held 1,093 patents.

bamboo from Japan was the best material. However, Edison was still not totally satisfied, and in November 1886 he sent agents to all parts of the globe looking for even better natural fibres for his lamps. In the end it was estimated that the Edison Lamp Company would "require a million and a half fibres at one cent a piece to pay for this expense".

Edison followed the same procedure to find the best phonograph record material when competition started to push him out of the market, and again when he took up X-rays shortly after Wilhelm Röntgen announced his discovery in December 1895. On that occasion Edison and his staff experimented with 150 different vacuum-tube designs over a two-month period, allowing him to produce considerably cheaper tubes. He and his staff also conducted experiments with nearly 1,300 compounds in order to find the best fluorescing salt for X-ray photographic plates and screens; this led to his calcium-tungstate fluoroscope. Near the end of

his career, in 1927, he joined forces with his friend Henry Ford to discover the most suitable alternative source of natural rubber in an attempt to free the United States from its reliance on this British monopoly.

The highlights of Edison's career are well known. After little formal education he became a telegraph engineer and roamed the country for work, while at the same time continuing a programme of self-improvement, in particular in chemistry and the electrical sciences as they related to telegraphy. He showed early signs of his entrepreneurial spirit in his keenness both for making money and for inventing. His early patents concern telegraphy. Among them were an improved stock-ticker, an automatic telegraph repeater and the duplex telegraph which, combined with the existing duplex method, enabled four messages to be transmitted over the same wire.

An interesting feature of his inventive process to which Israel draws our attention is how Edison exploited his knowledge in one field to develop what appear to be unrelated devices. Thus, his experience in telegraphy technology led him to invent an early copying system, the electric pen.

Edison's best-known inventions are the carbon transmitter for the telephone and the phonograph (gramophone). The former led to a dispute over who held the ownership of this invention in order to exploit the market. With the latter Edison had to learn the hard lesson that it is not always easy to develop a prototype into a marketable commodity.

Edison constructed the first central electric power station at Pearl Street, New York, but tried to hinder the development of alternating-current power systems by his competitors. To get the American public to believe that high-voltage a.c. power was dangerous, he advised the government that a humane form of execution could be achieved by means of the electric chair



Leading light: Edison's work on developing the light bulb led to electric lighting in homes.

powered by Westinghouse's a.c. generators.

By the end of his career Edison held 1,093 US patents. It can be argued, however, as becomes clear from Israel's book, that perhaps Edison's most original and significant invention was to establish the industrial research laboratory, his "invention-factory", at Menlo Park, and at his final home at West Orange.

The book dispels the popular myths surrounding Edison's laboratories. He was not the 'lone inventor', although this was a myth he was happy to foster. He appreciated the importance of research teams and employed men who had had a scientific education. It was his experience in trying to find such scientifically trained people for his laboratories that made him promote technical education.

For all Edison's success, Israel identifies several important weaknesses in his approach. Edison was good at developing a technically superior product but not at meeting the tastes of consumers. That is why ultimately he was not successful in spotting the trends in either the gramophone or the motion-picture business.

As scientific research became more specialized, Edison's approach seemed increasingly anachronistic. Early on in his career he had exclaimed that "There is as much difference between an inventor and a scientist as there is between an explorer and a geographer". For Edison the gathering of facts continued to take precedence over the development of theory. This approach had been successful when developing the telegraph but in the end was no longer viable. He was still following this approach in the 1920s when looking for new sources for natural rubber, while the real research was now the theoretical study of the molecule that was to lead to synthetic rubber. As Israel concludes, echoing the historian Thomas Hughes, the word 'inventor' had fallen into disuse by the time of Edison's

death, to be replaced by 'industrial scientist'. Israel's detailed analysis shows how 'the wizard of Menlo Park' was a seminal figure in this transformation. □

Willem Hackmann is at the Museum of the History of Science, University of Oxford, Broad Street, Oxford OX1 3AZ, UK.

Hypnotism and the rise of anaesthesia

Mesmerized: Powers of Mind in Victorian Britain

by Alison Winter

University of Chicago Press: 1998. 450 pp. \$30, £23.95

Ruth Richardson

This is one of those rare books that even a reviewer reads more slowly as the end approaches, to eke the experience out a little longer. It is a work of fine scholarship in the history of the sciences which, though richly packed with information and argument, reads as easily as a good novel. The publishers expect a rush of buyers: it is priced for a popular sale, well illustrated and features a sensational jacket design: a dramatic scene in which four late-Victorian male doctors are shown in rapt observation of a mesmerized woman in their laboratory.

Alison Winter's breadth of view and the diversity of her findings make for an unusual read. Her fundamental argument is that mesmerism was a catalyst that transformed medical and scientific authority in mid-nineteenth-century Britain.

Animal magnetism was discovered by the eighteenth-century doctor Franz Mesmer, who believed he was applying Newtonian philosophy to animal health by manipulating

the body's own magnetic field. The extraordinary vitality of interest in mesmerism just before the French Revolution in 1789 had been almost forgotten until the historian Robert Darnton published a book on the subject in 1968 (*Mesmerism and the End of the Enlightenment in France*; Harvard Univ. Press).

As with many other continental influences, it seems that the Revolutionary and Napoleonic Wars delayed interest in the 'new science' of mesmerism on the British side of the Channel until the 1830s. Thereafter it gained strength, first as a healing science, then manifesting itself in intuitive diagnostics, prescience, phrenopathy and anaesthesia. Other unexplained phenomena, such as electrobiology and table-turning, accrued to the movement before some of its vitality entered what we now regard as 'mainstream' sciences, like physics and medicine, and the remainder was absorbed by hypnotism and spiritualism after the middle of the century.

Both Darnton and Winter convey how open to new scientific phenomena society then was, and how undifferentiated its sciences, before the arrival of animal magnetism. For Darnton the craze for mesmeric phenomena was the end of the Enlightenment in France: belief in Reason succumbed to an onslaught of popular enthusiasm for pseudo-science. He followed the story of animal magnetism in France to its dissipation in the post-Revolutionary period. It may be coincidental that interest in Britain peaked just before Chartism in the revolutionary year of 1848, but for Winter mesmerism did not end the Enlightenment in Victorian Britain. Rather, what we regard as modern medicine and science developed in the way they have partly as a result of its impact. *Mesmerized* argues that, by rendering subjective evidence suspect, the investigation of mesmerism and its associated phenomena served to define evidence that excluded the subjective as 'scientific'.

The volume opens dramatically with a mesmeric duel at a London tea-party in November 1844, at which a strong-minded Jane Carlyle successfully withstood the powerful animal eyes of a distinguished magnetizer. Pervasive public curiosity in mesmerism's mysterious consciousness-altering effects induced forward-looking doctors to debate and investigate its phenomena by trying mesmerism on their patients, just as they tried other scientific discoveries such as the stethoscope.

Although animal magnetism produced cures in some cases, its unpredictability was uncongenial: some individuals would fall into a mesmeric trance within a few minutes; others might take hours, or indeed, never succumb at all. And even when under the influence, mesmerized subjects could behave in unexpected — sometimes quite subversive — ways, wreaking havoc on the



Under control: at first a matter of curiosity, mesmerism was later added to the doctor's medical armoury.

proprieties of the doctor–patient relationship, and on medical reputations.

Medical men were pressured to take a view: to investigate, explain or reject the new science. The latter course was eased for many when in 1838 Thomas Wakley (founder of *The Lancet*, and a keen exposé of quackery) proved beyond doubt before a large audience that the eminent professor John Elliotson of University College London had been hoaxed. Elliotson resigned. But despite *The Lancet's* ridicule, public and medical interest seems to have continued unabated.

Neither medicine nor science had a monopoly on investigation, as mesmerizing skills could be acquired relatively easily. Itinerant lecturers took their skills to public and private audiences all over the country. Doctors witnessed them, accepted tuition from and argued with them. Awareness of occasional hoaxes did not nullify the experiences of people who to their own mystification had themselves experienced or witnessed mesmeric phenomena. Professional disarray concerning the status of the evidence for and against mesmerism served to question not only the very nature of evidence but also the scientific credentials of doctors and scientists on all sides of the debate. Accusations of fraudulence were not one-way.

Mesmeric anaesthesia was at first a matter of curiosity in public performances of animal magnetism in which mesmerized subjects seemed oblivious to otherwise painful stimuli, such as being stuck with pins or burnt. Its use in surgery was a logical step, and when tried in 1842 in a case of leg amputation at the thigh, proved an instant success. Fierce controversy ensued. Not surprisingly, patients were keen to try it, and many doctors were keen to assist them. Others simply asserted that anaesthetized patients were fakers.

Winter argues convincingly that the deliberate suspension of pain during surgery was a by-product of mesmeric research, and that its success stimulated the development and the swift and widespread adoption of chemical anaesthesia. The notion that doctors had neglected their patients by failing to ease their pain, and that mesmerism could deliver an effective alternative, was such a spur to doctors suspicious of mesmerism that they disregarded the dangers and fatalities of ether and chloroform, to possess the new medical grail: painless — but 'scientific' — surgery.

This finding looks like a revolution in the history of anaesthesia. It is reinforced with two very different clusters of evidence. First, the leg amputation at the thigh performed by the famous surgeon Robert Liston to demonstrate the first use of ether anaesthesia in the United Kingdom took place in 1847 with triumphal symmetry in the same teaching theatre in which Elliotson had first displayed his mesmerized patients. Liston is

reported to have exclaimed, "this Yankee dodge beats mesmerism hollow". Second, there are the published data generated from a surgical production line developed by James Esdaile, a Scottish surgeon, between the mid-1840s and 1850s in the Native Hospital, Calcutta. Here, patients were prepared for the knife by a team of mesmerizing assistants. Sceptics had very properly insisted on reliable and repeatable evidence. But even where this was available (as it evidently was in Calcutta), mesmeric anaesthesia failed to attain 'scientific' credibility.

The problem seems partly to have been that unexplained phenomena that depended on what we call suggestibility was characterized by rationalists as suspect. Mesmerism was responsible for the development of the notion of suggestion, by James Braid, and of unconscious action, in the work of Michael Faraday in investigating table-turning by the inculcation of those whose joined hands involuntarily spun the table.

One of this book's strengths is that it conveys the strangeness — and the newness — of mesmerism, which so puzzled and intrigued contemporaries. Mesmerism's scientific and medical impacts represent only part of the story told in this extraordinary book, which examines its wide cultural repercussions in the arts (*A Christmas Carol*, *La Somnambula*), including the development of the role of the orchestral conductor, and in politics, the development of the notion of 'consensus'.

The word 'mesmerized' remains highly charged because of what we think we know of the power of the mesmerizer, and because, despite our current knowledge, many of its phenomena seem to remain beyond our understanding: dangerous, *unscientific*. Such reactions indicate the influence the nineteenth century continues to exert over our own intellectual culture, an influence this book brilliantly explains. □

Ruth Richardson is in the Department of Anatomy, University College, London WC1E 6BT, UK.

Mother love and its selective advantage

Maternal Effects as Adaptations

edited by Timothy A. Mousseau and Charles W. Fox

Oxford University Press: 1998. 348 pp.
£49.50, \$65

Steve Stearns

Mothers love their children, of that we can be sure. Some biologists, who love their mothers as much as anyone, interpret the effects that mothers have on their offspring as adaptations shaped by evolution.

A female parasitoid wasp determines the sex of her haplodiploid offspring by con-

trolling insemination: inseminated eggs become daughters. She lays daughters in large larvae, sons in relatively small ones — the reproductive success of daughters is more dependent on large body size than is that of sons.

A socially dominant red deer hind gives birth to more sons than daughters, probably by selective abortion of female fetuses. Her sons have a better chance of becoming dominant and fathering many grandchildren than do the sons of subordinate females. Subordinate females, on the other hand, have offspring with normal sex ratios. So are maternal effects usually adaptations? The case is clearer for these examples than it is for most of the effects mentioned in this book, and the authors admit as much.

Although there are many such examples, the editors and authors of the book complain that adaptive maternal effects have been neglected and that this must change. How do they arrive at the impression of neglect in the midst of plenty? To an evolutionary quantitative geneticist, taking an effect seriously means studying its genetics. This they do by extending the methods of quantitative genetics to include inter-generational effects of parental phenotypes on offspring phenotypes, and they succeed rather well.

Why did they think that the time was ripe to focus on maternal effects? Here I was less satisfied. Plant and animal breeders have tended to play down maternal effects, and the dominant texts on plant and animal breeding make little mention of them. The neglect has thus not been the fault of evolutionary biologists, but of non-evolutionary quantitative geneticists, who can hardly be blamed for not addressing problems they did not realize they were supposed to solve. I wonder whether one needs a whole book to make that simple point.

Some authors seemed to value complexity for its own sake. Other contributions were more satisfying. Roff's review of methods is clear, useful and concise. The reviews of maternal effects in flowering plants, insects, fish, amphibians and rodents are competent summaries of the state of the art. Denlinger's chapter on the transgenerational control of fly diapause shows how much can be achieved with classical techniques; here the way ahead appears to be molecular rather than quantitative genetics.

In brief, this multi-authored symposium volume has some new, creative contributions to methods, some useful reviews of the state of the art and a few chapters from people who should have done a better job. All the authors are from North America. I hope that this does not declare a prejudice about where significant work is being done. □

Steve Stearns is at the Zoology Institute, University of Basle, Rheinsprung 9, CH-4051 Basle, Switzerland.

Brown dwarfs: the stars that failed

C. G. Tinney

In recent years, new astronomical instruments have reversed three decades of fruitless searching for brown dwarfs, the failed stars that are too small to burn nuclear fuel. These discoveries have confirmed predictions about the importance of methane and dust in the atmospheres of brown dwarfs. But they also demonstrate that, contrary to expectation, brown dwarfs do not contribute significantly to our Galaxy's dark matter.

Most stars spend most of their lives in a state of pressure balance maintained between gravitational contraction and the energy generated by nuclear reactions. In 1963 Kumar¹ suggested that there may exist a class of star-like bodies, with masses too low to create the central temperatures and densities required to ignite nuclear fusion reactions. Although initially called black dwarfs, these 'failed stars' subsequently became known to astronomers as brown dwarfs². The lowest-mass stars can maintain a quasi-equilibrium luminosity for almost 6,000 Gyr (ref. 3), which is roughly 400 times the current age of the Universe. Brown dwarfs, in contrast, fade throughout their lifetime, cooling to temperatures below 1,000 K and becoming undetectable by direct observation after just a few billion years (refs 4–6). This has engendered considerable interest in their discovery, as they are a prime candidate for the dark matter which composes more than 90% of the Milky Way's mass⁷.

The simplest definition of a brown dwarf depends purely on mass, with theoretical models^{4–6} predicting a lower mass limit for fusion burning stars, with the same elemental mix as the Sun, of 0.07 solar masses ($0.07M_{\odot}$; equivalent to 74 jovian masses). However, one can imagine the formation of objects below this mass by both star formation processes (molecular-cloud fragmentation and collapse), and by planet formation processes (instabilities in, or accretion within, the disk around a young star⁸). Astronomers generally consider only the former to be brown dwarfs. This is not as arbitrary as it may seem, as only brown dwarfs formed like stars are interesting as dark matter, given that the mass of any planet will always be much less than that of its parent star.

The past four years have seen success finally achieved in the hunt for brown dwarfs. These detections have confirmed predictions that both methane and dust play an important role in determining the behaviour of these objects. In particular, they significantly affect the emergent spectrum in the broad bands used to seek brown dwarfs, and in the detailed spectra used to classify and study them. The detection of brown dwarfs in significant numbers, when combined with results for the space density of low-mass stars and gravitational microlensing results, allows us to conclude that brown dwarfs do not contribute to the dark-matter budget of our Galaxy.

The technical challenge

Figure 1 shows model results for low-mass stars and brown dwarfs. A clear difference is seen between stars, which within several hundred million years reach a constant luminosity (L) and temperature (T_{eff}), and brown dwarfs, which continually fade and cool. The lowest-mass stars are very faint— $5 \times 10^{-4}L_{\odot}$ (here $L_{\odot}$ indicates the solar luminosity). Brown dwarfs have even lower luminosities and get dimmer as they age. The first challenge, therefore, to detecting brown dwarfs has been this faintness. The second challenge has been their rarity. Although the number of objects of a given mass per volume of space increases with decreasing mass, the extreme faintness of brown dwarfs means that they can only be detected in the very near solar neighbourhood (60–300 light years). Thus maximizing the small search volume available requires a technically challenging level of wide-field imaging.

A number of developments have been responsible for the turn-

around after 30 years of unsuccessful effort^{9–12}. First, the size of commonly available optical and infrared detectors has increased in recent years by factors of between four and sixteen. This, in combination with the availability of the 10-m Keck telescopes, has made possible large surveys for brown dwarfs in nearby, young star clusters^{13,14}. Second, the commencement of two all-sky infrared surveys—2MASS¹⁵ and DENIS¹⁶—has dramatically increased the volume of the solar neighbourhood surveyed, producing the first of what promises to be a continuous stream of field brown-dwarf discoveries¹⁷. Third, the introduction into common use of powerful infrared spectrographs has permitted the discovered brown dwarfs to be studied in the wavelength range where most of their flux is emitted. Last, new high-precision radial-velocity techniques have produced numerous giant-planet discoveries, which seem to show a difference between the frequency of planet and brown-dwarf formation¹¹.

Methane and dust

Of the many theoretical predictions made before the discovery of brown dwarfs, those related to the importance of methane and dust stand out. In 1994 Tsuji *et al.*¹⁸ predicted that methane (CH_4) would play a dominant role in determining the emergent infrared spectrum of the coolest ($T_{\text{eff}} < 1,500$ K) brown dwarfs. In particular, it was predicted that in certain flux windows CH_4 would reverse the usual colours seen in brown dwarfs, making them blue instead of red. The possibility that dust might form in the atmospheres of the lowest-mass stars and brown dwarfs had also been perceived before the discovery of brown dwarfs, with models predicting warming of their upper atmospheres via a greenhouse effect (ref. 19), and possible changes in elemental abundances due to condensation²⁰.

Among the first searches for brown dwarfs to achieve success was that of Nakajima *et al.*²¹, which targeted the companions to nearby stars. In 1995, these authors announced the discovery of Gliese 229B—an object so faint²² ($L = 6.4 \times 10^{-6}L_{\odot}$) and so cool^{23,24} ($T_{\text{eff}} < 1,000$ K), that it was immediately accepted as a bona fide brown dwarf. Gliese 229B's extreme nature gives it properties unlike any other astronomical object. Its infrared spectrum contains strong CH_4 features, and it appears blue in the $J - K$ colour—spectacularly confirming Tsuji's predictions. The confirmation that brown dwarfs can appear blue in some colours has been important in the design of subsequent survey strategies, while the strength of the CH_4 seen in Gliese 229B immediately suggests that this signature be used to winnow such rare objects from large samples.

Confirmation of the presence of dust in brown dwarfs has been more problematic, largely because dust effects are less amenable to simple modelling. Had observers only known it, the elemental abundance changes produced by dust had been clear for almost ten years, in the form of the object GD165B²⁵. Before the discovery of Gliese 229B, GD165B was the coolest and lowest-luminosity ($T_{\text{eff}} = 1,900$ K, $L = 1.2 \times 10^{-4}L_{\odot}$) star-like object known^{26–28}. Its unusual optical spectrum, with none of the TiO and VO absorption which dominates the spectra of the lowest-mass stars, tended to result in it being sidelined as an 'oddball' object. This classification

had to be revised, however, with the discovery in 1997 of the first isolated brown dwarfs Kelu-1²⁹ and DENIS-PJ1228 – 1547³⁰. Both objects show spectra similar to GD165B (Fig. 2). Both are also brown dwarfs, as confirmed by the ‘lithium test’ (see below)^{31,32}. In combination, these results suggested that GD165B is simply what dwarf objects look like at $T_{\text{eff}} < 2,000$ K. This in turn provided the impetus to take a closer look at molecular equilibria. Subsequent detailed models³³ indicate that the molecules TiO and VO are indeed depleted at $T_{\text{eff}} < 2,000$ K by the condensation of solids containing Ti and V, so producing the observed spectral changes.

This depletion by dust formation has required the extension of the well established stellar OBAFGKM classification sequence with a proposed ‘L’-type, to include those objects cooler than $T_{\text{eff}} < 2,000$ K, which do not fit into the existing M-type. These L-dwarfs are characterized by spectra showing strong absorption bands of CrH (refs 27, 34), strong lines of the rare alkali metals Cs and Rb, and broad lines of alkali metals K and Na (ref. 32). All these features will play an important role in studying L-dwarfs, with the last being particularly important, as their strength depends critically on the extent of dust condensation and settling³⁴. The all-sky infrared surveys—in particular 2MASS—have played a vital

role in the process of understanding L-dwarfs by providing the first large samples of them for detailed study. In the absence of age information, L-dwarfs can not automatically be classified as brown dwarfs—though a significant fraction (7 of 23) of those discovered to date are¹⁷. As only the oldest and lowest-mass stars reach such temperatures²⁷, it appears that most L-dwarfs are also brown dwarfs.

Dust condensation will not only change the elemental abundances in brown dwarfs, but will also modify the emergent spectrum via scattering, absorption and re-radiation. Pre-discovery models by Tsuji *et al.*¹⁹ examined the effect on low-mass star and brown-dwarf atmospheres of a few selected dust species with a simple grain-size distribution. These models showed that dust should produce a greenhouse effect, warming the upper atmosphere and decreasing the strength of various molecular absorption bands. These predictions were subsequently verified for the lowest-mass stars¹⁹ and for GD165B²⁶. However, whereas these results have confirmed the general importance of dust formation, the details of the process remain unclear. In particular, as dust particles form they will undergo gravitational settling. Precise modelling of the settling rates requires an understanding of the specific species condensing and of their grain-size distribution. Models are only beginning to

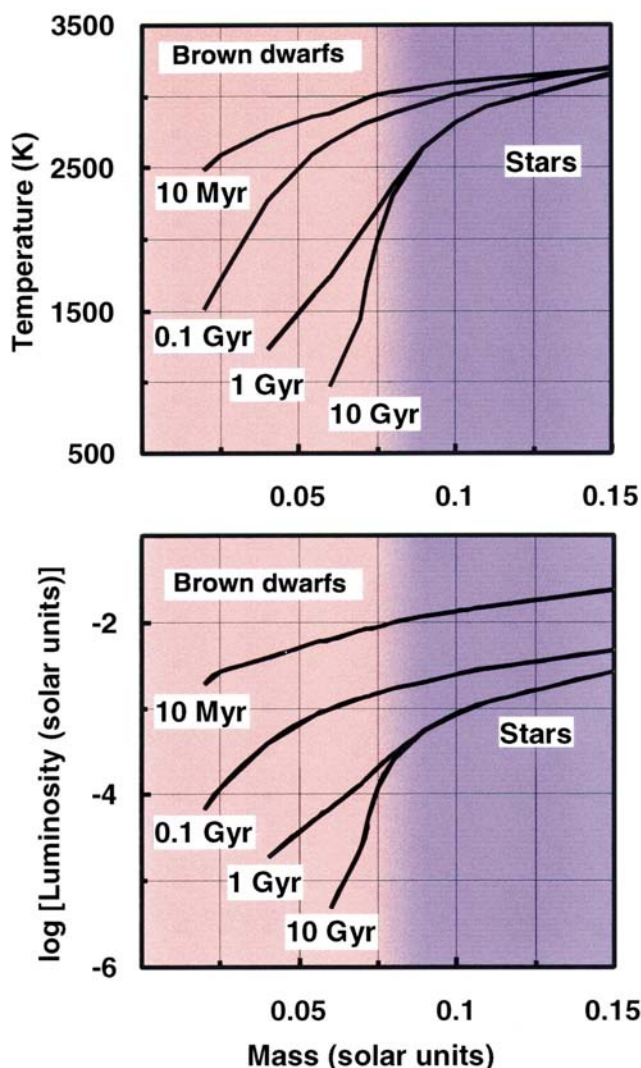


Figure 1 Luminosity and temperature as a function of mass and age for stars and brown dwarfs⁶. The shaded regions show the mass ranges of stars and brown dwarfs, separated by a transition region in which both gravitational contraction and nuclear burning provide energy. Note in particular that the properties of brown dwarfs depend strongly on age, whereas those of stars are essentially independent of age after a few hundred million years.

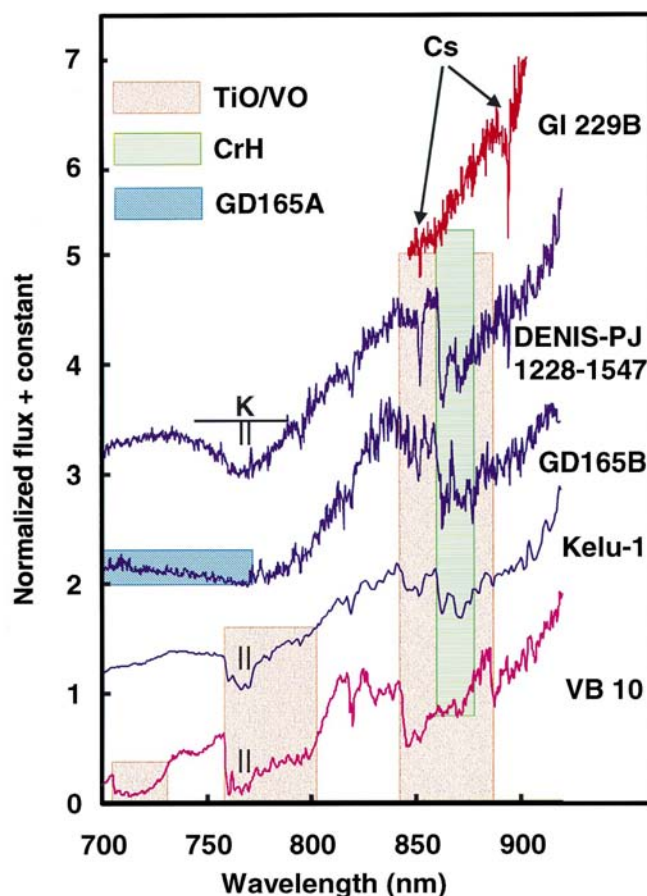


Figure 2 Sequence of optical spectra showing the effects of increasing temperature for the brown dwarfs Gliese 229B³⁶, DENIS-PJ 1228 – 1547³¹, Kelu-1³⁷ and the low-mass star VB10²⁸. The spectra are normalized at 880 nm, and offset by (in order) 4.5, 3.0, 2.0, 1.0 and 0.0. The shaded regions highlight the TiO and VO features typical of M-dwarfs (shown pink), and the typical L-dwarf CrH feature (shown green). Also marked are lines due to Cs and K, the latter becoming exceptionally strong in cooler L-dwarfs and producing the broad feature indicated in DENIS-PJ 1228 – 1547. The hatched blue region shows where the GD165B spectrum is contaminated by its companion.

include these effects, forcing a reliance at present on observations and simple assumptions.

Models which assume that dust remains suspended at the altitude of formation produce a poor match to the optical spectra of GD165B-like L-dwarfs, with models which simulate instantaneous settling doing better³⁴. This suggests either the formation of dust clouds which cover only a small fraction of the brown-dwarf's surface, or that dust is settling to levels below the photosphere (the depth in the atmosphere from which photons escape to be detected). Gliese 229B is a powerful test, because at its lower temperature dust should form even more readily than in L-dwarfs²⁰. In this case, infrared spectra argue strongly against suspended-dust models²⁶. Optical spectra^{35,36} indicate the presence of some dust, though only a small quantity seems to be implied³⁷. These results indicate that neither of these simple assumptions—that all dust settles rapidly, or that dust never settles—describe reality. The truth must lie somewhere in between. Future studies, including the growth and settling of each grain species, will be required in order to understand the grain–height, grain–size and grain–density distributions, and their effect on the emergent spectrum. In particular, these models will be vital to determining whether cloud formation and meteorology is as important in brown dwarfs as we know it to be in giant planets.

Weighing and counting brown dwarfs

The most pressing question facing brown-dwarf research is “Does star formation produce enough brown dwarfs to contribute significantly to dark matter?”. Answering this requires both counting brown dwarfs and estimating their masses. Unfortunately, the latter remains problematic. Direct measurements can only be made in a very small number of binary configurations. Few brown dwarfs in the requisite systems are yet known, and none have produced dynamical mass measurements. Currently, therefore, masses must be estimated by using measured L or T_{eff} values and the results of theoretical models. But such a conversion requires knowledge of each brown dwarf's age (see Fig. 1). With few exceptions, such information is impossible to obtain for brown dwarfs found in the general field of stars around the Sun.

However, we can get some information on the space density as a function of mass for field brown dwarfs (the mass function) by looking at the value for low-mass stars just above the brown-dwarf limit. Results from such studies are generally parametrized in terms of a power-law fit to the observed mass function. Using the form $\psi(m) \propto m^{-\alpha}$, where ψ is the number of objects per cubic parsec per unit mass interval, m is mass, and α is the power-law index, an index of $\alpha \approx 2.3$ is seen in stars more massive than $1M_{\odot}$ (ref. 38). Such a mass function, if extrapolated into the brown-dwarf regime, with reasonable assumptions being made about the minimum mass for brown-dwarf formation⁸, can predict a significant quantity of dark matter in the form of brown dwarfs. But numerous studies of the mass function for the lowest-mass stars ($M = (0.1\text{--}0.7)M_{\odot}$) have been performed⁹ and their general result is that at these masses a shallower index of $0.7 < \alpha < 1.8$ (refs 39–41) is observed. This is not steep enough to produce significant dark matter in the form of brown dwarfs. It should be stressed that power-law parametrizations are adopted for descriptive simplicity, rather than because they reflect a detailed understanding of star-formation processes. Indeed, it has been suggested that a log-normal parametrization ($\log(\psi(\log m)) \propto (\log[m/m_c])^2/(2\sigma^2)$, where ψ and m are as for the power law, and m_c and σ are the log-normal parameters) may better describe the way in which the distribution of parameters affecting star formation maps into the mass function⁴². The acceptance of such a hypothesis must await the clear detection of a fall in the mass function in the brown-dwarf regime. In any case, a log-normal parametrization integrates to less dark matter than a power law. The available data make the presence

of significant brown-dwarf dark matter in the solar neighbourhood impossible.

The advantages of youth

An alternative technique is to measure the mass function in young (<300 Myr) star clusters of known age. This has the added advantage of probing brown dwarfs when they are at their brightest. Such studies have been productive, and produced the first accepted brown-dwarf detections^{43,44}, with the backbone of these searches being the ‘lithium test’. Li is a fragile element, which is destroyed by nuclear reactions at temperatures above 2.4×10^6 K. This is below the minimum central temperature required to sustain hydrogen-burning, so all stars can also burn Li (ref. 45). Moreover, because low-mass stars and brown dwarfs are fully convective, they cycle material from their surface to their core. As a result, an object at the $0.07M_{\odot}$ minimum stellar mass will destroy 99% of its Li within only 2.2×10^8 yr, while brown dwarfs below $0.06M_{\odot}$ will take more than 10^{10} yr to destroy a similar Li fraction⁶. Searches for the 670.8-nm Li line therefore provide direct mass information, as its detection provides a firm upper mass limit of $0.06M_{\odot}$.

This technique has proved a number of field objects to be brown dwarfs^{17,29,31,32,46,47}—a list which will grow as more and more L-dwarfs are discovered. More importantly, the lithium test has confirmed numerous brown dwarfs in star clusters^{43–45}, demonstrating that their mass functions extend below the hydrogen burning limit. Large Pleiades surveys^{13,48} find a power-law index $\alpha < 1.5$ at masses in the brown-dwarf regime. Similar results have been found for the stars and brown dwarfs in the much younger (<5 Myr) star-forming regions of Ophiucus, Taurus and Orion, where a mass function of index of $\alpha \approx 1.2$ is seen^{49,50}. The older Praesepe cluster has also revealed a shallow mass function for its member stars^{51,52}. Both in young clusters and in the solar neighbourhood, therefore, we find that star formation does not produce a significant dark mass in the form of brown dwarfs.

A spin-off from such studies is the luminosity¹⁴ at which the dividing line is seen between brown dwarfs showing Li and stars with no Li. This ‘lithium luminosity’ corresponds to a particular mass for the particular age of the cluster in question, and moreover, is predicted with much less model dependence than existing techniques for determining the ages of clusters. This has produced a robust determination¹⁴ of the Pleiades' age as 125 ± 10 Myr, and offers the prospect of precise age determinations for a large number of young open clusters.

Missing mass . . . still missing

Dynamical studies of the rotation of stars and gas within the disk of our Galaxy indicate that the disk is embedded within a massive dark halo, which dominates the Galaxy's mass budget⁷. Similar studies in other spiral galaxies indicate this configuration is common⁵³. It has long been considered possible that there exists a population of dark halo objects with a mass function dominated by brown dwarfs that formed early in the Galaxy's history. Determining the mass function in the dark halo is therefore of fundamental importance. Unfortunately, other than via gravitational microlensing (see below), no dark halo objects have ever been detected, let alone had their mass function measured.

Nevertheless, we can gain some insight into early Galactic star formation by studying members of the old (10–15 Gyr) globular clusters of the spheroidal population. These are compact condensations of ancient stars, and have long been targets for mass-function studies. Brown dwarfs in such old populations are inaccessible, having faded to near invisibility ($L < 10^{-5}L_{\odot}$). However, as for the disk, headway can be made by looking at the stellar mass function just above the brown-dwarf limit. Such studies have received new impetus with the commissioning of the refurbished imaging instruments on the Hubble Space Telescope, which are used to separate the stars in crowded globular cluster environments. As an

example, such observations⁵⁴ have detected members of the cluster NGC6397 that are as faint as $6.3 \times 10^{-4} L_{\odot}$. This corresponds to a mass of $0.09 M_{\odot}$, which is just $0.007 M_{\odot}$ above the cluster's minimum stellar mass. Multiple clusters have now been surveyed for low-mass stars, and although there is some variation in the mass function observed from cluster to cluster, power-law fits^{54–57} are consistently producing slopes of $0 < \alpha < 1.5$.

Last, we can look to the only observations which probe the dark halo directly—gravitational microlensing. The principles used in such searches are well reviewed elsewhere⁵⁸. Early results were interpreted as providing evidence for a dark halo at least partially composed of objects of brown-dwarf mass. However, the added statistics included in more recent work⁵⁹ indicate that, if the observed events are due to members of a standard isothermal dark halo, then $\sim 50\%$ of the halo's mass is in objects with preferred masses of $0.5^{+0.3}_{-0.2} M_{\odot}$. This is somewhat larger than the mass of a brown dwarf. Unfortunately, although the total halo mass observed is well constrained, the preferred mass is quite model dependent⁵⁹; this makes it problematic to determine the preferred mass. (Other interpretations also exist—see ref. 60 and references therein.) What can be said is that only a small subset of the likely halo models permit a significant dark halo mass in brown dwarfs⁵⁹, and that microlensing results strongly argue against a dark halo composed of planetary-mass objects⁶¹. Those models which do permit a brown-dwarf halo require a mass function strongly peaked at brown-dwarf masses, which seems hard to reconcile with the fact that every other observed population shows a shallow mass function for both the lowest-mass stars and brown dwarfs.

Implications

The actual detection of brown dwarfs has not only highlighted those properties of brown dwarfs which were predicted and subsequently confirmed (such as the importance of methane and dust), but has also emphasized those areas of interior and atmospheric theory which will require further study. In particular, the growth and settling of dust grains in low-temperature atmospheres, and the effect of the 'weather' in these atmospheres on brown-dwarf properties, require intensive investigation. The detections have also shown that star-formation processes, although they may produce large numbers of brown dwarfs, do not produce enough to have a significant effect on issues related to missing mass. No matter how nicely brown dwarfs would solve the baryonic dark-matter problem, it appears we must look elsewhere for a solution to this long-standing astronomical quandary. □

The author is at the Anglo-Australian Observatory, PO Box 296, Epping NSW 1710, Australia.

- Kumar, S. S. The structure of stars of very low mass. *Astrophys. J.* **137**, 1121–1125 (1963).
- Tarter, J. C. in *Astrophysics of Brown Dwarfs* (eds Kafatos, M. C., Harrington, R. S. & Maran, S. P.) 121–138 (Cambridge Univ. Press, 1986).
- Laughlin, G., Adams, F. & Bodenheimer, P. in *Very Low-mass Stars and Brown Dwarfs in Stellar Clusters and Associations* (Cambridge Univ. Press, in the press).
- Burrows, A., Hubbard, W. B. & Lunine, J. I. Theoretical models of very low mass stars and brown dwarfs. *Astrophys. J.* **345**, 939–958 (1989).
- D'Antona, F. & Mazzitelli, I. New pre-main-sequence tracks for M less than or equal to 2.5 solar mass as tests of opacities and convection model. *Astrophys. J. Suppl. Ser.* **90**, 467–500 (1994).
- Chabrier, G., Baraffe, I. & Plez, B. Mass-luminosity relationship and lithium depletion for very low mass stars. *Astrophys. J.* **459**, L91–L94 (1996).
- Freeman, K. in *Unsolved Problems of the Milky Way* (eds Blitz, L. & Teuben, P.) 645–650 (IAU Symp. 169, Kluwer, Dordrecht, 1996).
- Bodenheimer, P. in *Brown Dwarfs and Extrasolar Planets* (eds Rebolo, R., Martín, E. L. & Zapatero Osorio, M. R.) 115–127 (ASP Conf. Ser. 134, Astron. Soc. Pacific, San Francisco, 1998).
- Reid, N. in *The Bottom of the Main Sequence—And Beyond* (ed. Tinney, C.) 53–62 (Springer, Heidelberg, 1995).
- Henry, T. J. in *Brown Dwarfs and Extrasolar Planets* (eds Rebolo, R., Martín, E. L. & Zapatero Osorio, M. R.) 28–35 (ASP Conf. Ser. 134, Astron. Soc. Pacific, San Francisco, 1998).
- Marcy, G. & Butler, R. P. Detection of extrasolar giant planets. *Amnu. Rev. Astron. Astrophys.* **36**, 56–96 (1998).
- Hambly, N. in *Brown Dwarfs and Extrasolar Planets* (eds Rebolo, R., Martín, E. L. & Zapatero Osorio, M. R.) 11–19 (ASP Conf. Ser. 134, Astron. Soc. Pacific, San Francisco, 1998).
- Bouvier, J. et al. Brown dwarfs and very low-mass stars in the Pleiades cluster: a deep wide-field imaging survey. *Astron. Astrophys.* **336**, 490–502 (1998).
- Stauffer, J., Schultz, G. & Kirkpatrick, J. D. Keck spectra of Pleiades brown dwarf candidates and a precise determination of the lithium depletion edge in the Pleiades. *Astrophys. J.* **499**, L199–L203 (1998).
- Skrutskie, M. F. et al. in *The Impact of Large Scale Near-IR Sky Surveys* (eds Garzón, F. et al.) 25–32 (Kluwer, Dordrecht, 1997).

- Epchtein, N. in *The Impact of Large Scale Near-IR Sky Surveys* (eds Garzón, F. et al.) 15–24 (Kluwer, Dordrecht, 1997).
- Kirkpatrick, J. D. et al. Brown dwarfs discovered by 2MASS and the definition of a new spectral type cooler than "M". *Bull. Am. Astron. Soc.* **192**, 5504–5504 (1998).
- Tsuji, T., Ohnaka, K. & Aoki, W. in *The Bottom of the Main Sequence—And Beyond* (ed. Tinney, C.) 45–49 (Springer, Heidelberg, 1995).
- Tsuji, T., Ohnaka, K. & Aoki, W. Dust formation in stellar photospheres: a case of very low mass stars and a possible resolution on the effective temperature scale of M dwarfs. *Astron. Astrophys.* **305**, L1–L4 (1996).
- Sharp, C. M. & Huebner, W. F. Molecular equilibrium with condensation. *Astrophys. J. Suppl. Ser.* **72**, 417–431 (1990).
- Nakajima, I. et al. Discovery of a cool brown dwarf. *Nature* **378**, 463–465 (1996).
- Matthews, K. et al. Spectral energy distribution and bolometric luminosity of the cool brown dwarf Gliese 229B. *Astron. J.* **112**, 1678–1682 (1996).
- Allard, F., Hauschildt, P. H., Baraffe, I. & Chabrier, G. Synthetic spectra and mass determination of the brown dwarf GL229B. *Astrophys. J.* **465**, L123–L127 (1996).
- Oppenheimer, B. R., Kulkarni, S. R., Matthews, K. & Nakajima, T. Infrared spectrum of the cool brown dwarf GL229B. *Science* **270**, 1478–1479 (1995).
- Becklin, E. E. & Zuckerman, B. A low-temperature companion to a white dwarf star. *Nature* **336**, 656–658 (1988).
- Tsuji, T., Ohnaka, K. & Aoki, W. Evolution of dusty photospheres through red to brown dwarfs: how dust forms in very low mass objects. *Astron. Astrophys.* **308**, L29–L32 (1996).
- Kirkpatrick, J. D. et al. An improved optical spectrum and new model fits for the likely brown dwarf GD 165B. *Astrophys. J.* (submitted).
- Tinney, C. G., Mould, J. R. & Reid, I. N. The faintest stars—infra-red photometry, spectra and bolometric magnitudes. *Astron. J.* **105**, 1045–1059 (1993).
- Ruiz, M. T., Leggett, S. K. & Allard, F. Kelu-1: a free-floating brown dwarf in the solar neighborhood. *Astrophys. J.* **491**, L107–L110 (1997).
- Delfosse, X. et al. Field brown dwarfs found by DENIS. *Astron. Astrophys.* **327**, L25–L28 (1997).
- Tinney, C. G., Delfosse, X. & Forveille, T. DENIS-P J1228.2-1547—a new benchmark brown dwarf. *Astrophys. J.* **490**, L95–L98 (1997).
- Martin, E. L., Basri, G., Delfosse, X. & Forveille, T. Keck HIRES spectra of the brown dwarf DENIS-P J1228.2-1547. *Astron. Astrophys.* **327**, L29–L32 (1997).
- Allard, F., Hauschildt, P., Alexander, D. R. & Starrfield, S. *Annu. Rev. Astron. Astrophys.* **35**, 137–178 (1997).
- Tinney, C. G., Delfosse, X., Forveille, T. & Allard, F. Optical spectroscopy of DENIS mini-survey brown dwarf candidates. *Astron. Astrophys.* **338**, 1066–1072 (1998).
- Oppenheimer, B. R., Kulkarni, S. R., Matthews, K. & van Kerkwijk, M. The spectrum of the brown dwarf Gliese 229B. *Astrophys. J.* **502**, 932–943 (1998).
- Schultz, A. B. et al. First results from the space telescope imaging spectrograph: optical spectra of Gliese 229B. *Astrophys. J.* **492**, L181–L184 (1998).
- Burrows, A. & Sharp, C. M. Chemical equilibrium abundances in brown dwarf and extrasolar giant planet atmospheres. *Astrophys. J.* (submitted); also as preprint astro-ph/9807055 at (<http://xxx.lanl.gov>) (1998).
- Scalo, J. M. The stellar initial mass function. *Fund. Cosmic Phys.* **11**, 1–278 (1986).
- Reid, I. N. & Gizis, J. E. Low-mass binaries and the stellar luminosity function. *Astron. J.* **113**, 2246–2269 (1997).
- Kroupa, P. in *Brown Dwarfs and Extrasolar Planets* (eds Rebolo, R., Martín, E. L. & Zapatero Osorio, M. R.) 483–494 (ASP Conf. Ser. 134, Astron. Soc. Pacific, San Francisco, 1998).
- Tinney, C. G. The faintest stars: the luminosity and mass functions at the bottom of the main sequence. *Astrophys. J.* **445**, 1017–1018 (1995).
- Adams, F. in *Brown Dwarfs and Extrasolar Planets* (eds Rebolo, R., Martín, E. L. & Zapatero Osorio, M. R.) 3–10 (ASP Conf. Ser. 134, Astron. Soc. Pacific, San Francisco, 1998).
- Basri, G., Marcy, G. W. & Graham, R. J. Lithium in brown dwarf candidates: the mass and age of the faintest Pleiades stars. *Astrophys. J.* **458**, 600–609 (1996).
- Rebolo, R., Zapatero Osorio, M. R. & Martín, E. L. Discovery of a brown dwarf in the Pleiades star cluster. *Nature* **377**, 129–131 (1995).
- Magazzu, A., Martín, E. L. & Rebolo, R. A spectroscopic test for substellar objects. *Astrophys. J.* **404**, L17–L20 (1993).
- Tinney, C. G. The intermediate-age brown dwarf LP944-20. *Mon. Not. R. Astron. Soc.* **296**, L42–L44 (1998).
- Thackrah, A., Jones, H. & Hawkins, M. Lithium detection in a field brown dwarf candidate. *Mon. Not. R. Astron. Soc.* **284**, 507–512 (1997).
- Martin, E. L., Zapatero-Osorio, M. R. & Rebolo, R. in *Brown Dwarfs and Extrasolar Planets* (eds Rebolo, R., Martín, E. L. & Zapatero Osorio, M. R.) 507–514 (ASP Conf. Ser. 134, Astron. Soc. Pacific, San Francisco, 1998).
- Luhman, K. L. & Rieke, G. H. The low mass initial mass function in young clusters: L1495E. *Astrophys. J.* **497**, 354–369 (1998).
- Comeron, F., Rieke, G. H. & Rieke, M. J. Properties of low-mass objects in NGC2024. *Astrophys. J.* **473**, 294–303 (1996).
- Hambly, N. C., Steele, I. A., Hawkins, M. R. S. & Jameson, R. F. The very low-mass main sequence in the Galactic cluster Praesepe. *Mon. Not. R. Astron. Soc.* **273**, 505–512 (1995).
- Williams, D. M., Rieke, G. H. & Stauffer, J. R. The stellar mass function of Praesepe. *Astrophys. J.* **445**, 359–366 (1995).
- Ashman, K. M. Dark matter in galaxies. *Publ. Astron. Soc. Pacif.* **104**, 1109–1138 (1992).
- King, I. R., Anderson, J., Cool, A. M. & Piotto, G. The luminosity function of the globular cluster NGC 6397 near the limit of hydrogen burning. *Astrophys. J.* **492**, L37–L40 (1998).
- Piotto, G., Cool, A. M. & King, I. R. A comparison of deep HST luminosity functions of four globular clusters. *Astron. J.* **113**, 1345–1352 (1997).
- Pulone, L., De Marchi, G., Paresce, F. & Allard, F. The lower main sequence of omega Centauri from deep bubble space telescope NICMOS near-IR observations. *Astrophys. J.* **492**, L41–L44 (1998).
- Chabrier, G. & Mera, D. Determination of the globular cluster and halo stellar mass functions and stellar and brown dwarf densities. *Astron. Astrophys.* **328**, 83–94 (1997).
- Paczynski, B. Gravitational microlensing in the local group. *Annu. Rev. Astron. Astrophys.* **34**, 419–460 (1996).
- Alcock, C. et al. The MACHO project Large Magellanic Cloud microlensing results from the first two years and the nature of the galactic dark halo. *Astrophys. J.* **486**, 697–726 (1997).
- Zaritsky, D. & Lin, D. N. C. Evidence for an intervening stellar population toward the Large Magellanic Cloud. *Astron. J.* **114**, 2545–2555 (1997).
- Alcock, C. et al. EROS and MACHO combined limits on planetary-mass dark matter in the Galactic halo. *Astrophys. J.* **499**, L9–L12 (1998).

Acknowledgements. I thank J. Hawthorn, P. Butler and H. Sim for comments, and G. Chabrier, I. Baraffe, B. Oppenheimer & J. D. Kirkpatrick for material used in the figures.

A giant periodic flare from the soft γ -ray repeater SGR1900+14

K. Hurley*, T. Cline†, E. Mazets‡, S. Barthelmy†, P. Butterworth†, F. Marshall†, D. Palmer†, R. Aptekar‡, S. Golenetskii‡, V. Il'Inskii‡, D. Frederiks‡, J. McTiernan*, R. Gold§ & J. Trombka†

* UC Berkeley Space Sciences Laboratory, Berkeley, California 94720-7450, USA

† NASA Goddard Space Flight Center, Code 661, Greenbelt, Maryland 20771, USA

‡ Ioffe Physical-Technical Institute, St Petersburg 194021, Russia

§ The Johns Hopkins University, Applied Physics Laboratory, Laurel, Maryland 20723, USA

Soft γ -ray repeaters are transient sources of high-energy photons; they emit sporadic and short (about 0.1 s) bursts of 'soft' γ -rays during periods of activity, which are often broken by long stretches of quiescence. These objects are associated with neutron stars in young supernova remnants¹. The event of 5 March 1979 was the most intense burst to date, and the only one that showed a clear periodicity in the signal^{2,3}. Here we report the detection, on 27 August 1998, of an even more intense burst from a different soft γ -ray repeater. This event was characterized by 'hard' γ -rays at its peak, followed by a tail 300 s long with a soft spectrum and a clear periodicity of 5.16 s. The burst was probably initiated by a massive disruption of the crust of the neutron star, followed by an outflow of energetic particles rotating with the period of the star. A comparison of the events of 27 August 1998 and 5 March 1979 supports the idea that magnetic energy plays an important role in the genesis of such events. Although these giant flares are rare, they are not unique events and may occur at any time in a neutron star's activity cycle.

Four soft γ -ray repeaters (SGRs) are known. All appear to be associated with radio supernova remnants, indicating that they are young⁴ (<20,000 yr). SGRs are probably strongly magnetized neutron stars ('magnetars'⁵), in which, unlike the radio pulsars, the magnetic energy dominates the rotational energy. SGR0525-66 produced the unusual, energetic and periodic burst of 5 March 1979 (refs 3, 6, 7) and a series of subsequent, much smaller bursts^{8,9}. It lies towards the N49 supernova remnant in the Large Magellanic Cloud^{10,11}. A quiescent soft X-ray source has been identified that may be the neutron star¹². SGR1900+14, first detected in 1979, was, until recently, the least prolific SGR^{13,14}, hindering attempts to locate it precisely. Several lines of evidence suggested that it was associated with the Galactic supernova remnant G42.8+0.6 (ref. 15) and a quiescent soft X-ray source¹⁶. This possible association was strengthened by a source location obtained with the network synthesis method¹⁷, and more recently by triangulation¹⁸⁻²⁰, although because this X-ray source lies outside the remnant, the connection between the two could still be considered to be unresolved.

An observation of the quiescent soft X-ray source possibly associated with SGR1900+14 by the ASCA spacecraft in April 1998 showed that the X-rays exhibited a 5.16-s period²¹. In May of that year, SGR1900+14 came out of a long dormant phase, emitting strong, frequent bursts^{18,22}. On 27 August 1998, it emitted the exceptionally intense giant flare reported here, detected by instruments on the GGS-Wind²³, Ulysses²⁰, Rossi X-Ray timing Explorer²⁴ (RXTE), Beppo SAX, and Near Earth Asteroid Rendezvous (NEAR) spacecraft. The entire event profile is shown in Fig. 1 with Ulysses data at 0.5 s resolution. In very general terms, the burst rose to a maximum and decayed roughly as a power law in time with an index of about -1.8. However, the event onset is complex; Konus-

Wind observations (Konus is an experiment aboard the GGS-Wind spacecraft) resolve components <4 ms long. A sinusoidal component dramatically modulated the later part of the profile with varying amplitudes for the duration of the observation, the first direct detection of the 5.16-s periodicity at hard X-ray energies. Figure 1, inset, shows Ulysses data with 31.25-ms time resolution, demonstrating that the 5.16-s pulsations commenced ~35 s after the peak. It is clear that the pulse profile is considerably more complex than a single sinusoidal curve, with at least four maxima and minima in a single cycle.

A remarkable coincidence—the initiation of NEAR γ -ray monitoring only days before 27 August but after many months of silent cruise towards Eros—made possible the high-precision source localization of this event by triangulation; this was done by analysis of the arrival times at Ulysses, GGS-Wind, RXTE and NEAR. This is the only time, other than for the 5 March 1979 event^{10,11}, that an SGR has been localized by triangulation at three or more widely separated spacecraft, leading directly to an error box. All six source error annuli determined from the various two-spacecraft comparisons, are consistent with the coordinates of the quiescent soft X-ray source^{17,20} (J2000 right ascension 19 h 0.7 min 14 s, declination 9° 19' 19"). The details will be reported elsewhere, but we note that this positional agreement, as well as the agreement between the periodicities found in soft X-rays and in the giant-flare light curve, now leave no doubt about the association between the SGR and the quiescent X-ray source.

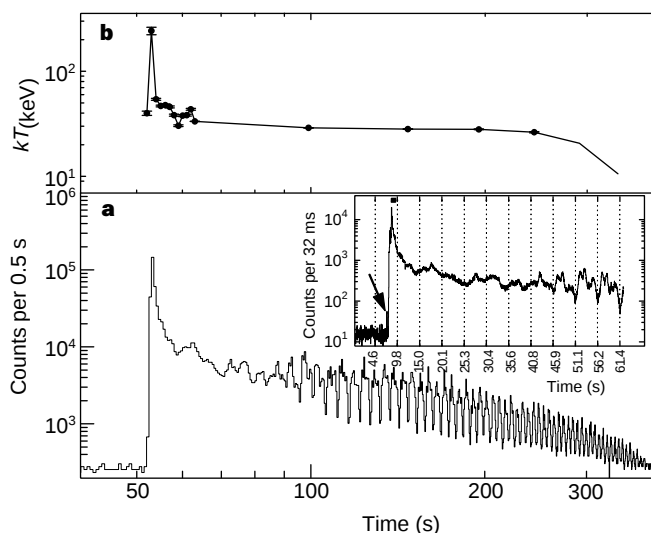


Figure 1 Ulysses data for the 27 August 1998 giant flare. **a**, 25–150 keV time history, corrected for dead-time effects, from the 0.5-s resolution continuously available real-time data. Zero seconds corresponds to 37,283.12 s UT at Earth. This event was so intense that it temporarily saturated or shut down some experiments, but because of the relatively small detection area of the Ulysses²⁸ sensor (20 cm²), it was not subject to severe dead-time or pulse pile-up problems; in fact solar flare data producing considerably higher count rates have been successfully analyzed with this instrument. Inset, 0.03125-s-resolution time history of the event from the triggered data, available for 64 s. The burst triggered on the precursor (arrow) ~0.4 s before the main peak. A grid is drawn to indicate the 5.16-s periodicity, showing its absence for the first ~35 s after the main peak. The short horizontal line at the top indicates the position of the hard spectral peak measured by Ulysses. Zero seconds corresponds to 37,327.81 s UT at Earth. **b**, Spectral temperature as a function of time. The spectra were measured by Ulysses in intervals with increasing durations of 1–48 s. No simple, two-parameter fit describes the spectrum well, in part because the measurement uncertainties are dominated by systematic effects. However, we have used an optically thin thermal bremsstrahlung spectrum to characterize approximately the spectral temperature.

Table 1 Comparison of the two bursts

Property	Burst	
	27 August 1998	5 March 1979
Rise time	Complex, structures <4 ms	Simple, <2 ms
Morphology of main peak	Complex structure, duration ~1 s	Complex structure duration ~150 ms ²⁶
Periodicity	5.16 s	8.1 s
Peak flux (erg cm ⁻² s ⁻¹)	$\geq 3.4 \times 10^{-3}$, >25 keV	$\sim 1.5 \times 10^{-3}$, >50 keV
Fluence (erg cm ⁻²)	$\geq 7 \times 10^{-3}$	$\sim 2 \times 10^{-3}$
Spectrum at peak, kT (keV)	240 (average over 1 s)	246 (average over 200 ms) ²⁷
Highest photon energy in peak	2 MeV	>1 MeV
Spectrum of pulsations, kT (keV)	30	30
Source distance (kpc)	~7 (G42.8+0.6)	~50 (N49)
Peak source luminosity (erg s ⁻¹)	$\geq 2 \times 10^{43}$	$\sim 5 \times 10^{44}$
Precursor observed?	Yes	No
Delay between main peak and periodic emission	35 s	None
Ratio of energy in main peak to total energy in burst	0.46	0.25
Source activity in months preceding the burst	Intense	None observed

The temperature of the energy spectrum of this event is shown in Fig. 1. With the exception of the peak, the temperature is $kT \approx 30$ keV, which is similar to SGR bursts in general. At the peak, however, the temperature averaged over a 1-s interval is $kT \approx 240$ keV. Measurements with finer time resolution were recorded by Konus-Wind, indicating a peak temperature of $\sim 1,200$ keV, and a maximum photon energy of 2 MeV. Hard spectra such as these are not characteristic of SGR bursts; one was observed for the peak of the 5 March 1979 event^{6,25}. Table 1 compares the properties of these two giant flares.

Comparisons between very intense bursts observed by different instruments are subject to numerous uncertainties. Dead-time effects, different time resolutions and energy ranges, and pulse pile-up are difficult or even impossible to correct for; hence the 'approximate' and 'greater than' symbols in Table 1. However, within these uncertainties, the parameters of the 27 August 1998 event are consistent with its having the largest peak flux and fluence of any of the several thousand SGRs and cosmic γ -ray bursts observed to date.

It has been suggested recently^{22,25} that the neutron stars associated with SGRs are magnetars, that is, that they have magnetic fields of several times 10^{14} G (ref. 5). This is based on observations of the quiescent counterparts in X-rays, which display pulsations with a slowly lengthening period; the spindown is interpreted as being due to magnetic dipole radiation. In the magnetar model, the giant flares of 27 August and 5 March are due to a readjustment of the magnetic field, accompanied by a massive, large-scale cracking of the neutron-star crust. In both cases, the initial hard spectrum would be produced by the conversion of magnetic energy to energy in a clean electron–positron and photon fireball uncontaminated by ions, which would soften the spectrum. The highest-energy photons observed are only slightly above the electron–positron pair production threshold; this is consistent with attenuation due to this process, although there is at present no direct evidence for a cut-off. Expanding away from the stellar surface, part of the fireball would be trapped in the magnetosphere, producing the observed soft-spectrum tails. The periodicity indicates that this emission was either anisotropic and/or that it occurred close enough to the neutron star to be occulted by it; the decay in intensity with approximately constant spectral temperature is interpreted as a shrinking in the volume of the emission region. The complex pulse structure implies that several regions of the magnetosphere were involved. We note that, despite the factor of 25 difference between

the peak luminosities of the 27 August and 5 March events, the ratios of peak to total energy are within a factor of 2 of each other, suggesting that similar magnetic field geometries may be important. As the soft spectrum that follows the intense main peak in both cases is attributed to radiation from an optically thick pair plasma trapped in the neutron star's magnetosphere, the magnetic field strength may be estimated from the energy in this component⁵:

$$B > 4 \times 10^{14} \left(\frac{\Delta R}{10 \text{ km}} \right)^{-3/2} \left(\frac{1 + \Delta R/R}{2} \right)^3 \left(\frac{E_{\text{tail}}}{3.6 \times 10^{44} \text{ erg}} \right)^{1/2} \text{ G}$$

where R is the radius of the neutron star and ΔR (~ 10 km) is the outer radius of the magnetic flux loop containing the pair plasma. For the 5 March event, this gives $B > 4 \times 10^{14}$ G; for the 27 August event, $B > 10^{14}$ G, providing a confirmation of the magnetar model which is independent of the observation and interpretation of the spindown, but consistent with it.

The existence of a strong magnetic field helps to explain the high luminosities encountered in both events, five to six orders of magnitude greater than the Eddington limit. A strong magnetic field suppresses the Compton scattering cross-section, and reduces the opacity⁵.

The giant flare of 5 March 1979 was observed to precede the much smaller event series from SGR0525–66. Observations during the preceding six months failed to reveal any source activity, and it was speculated at the time that this was a unique, catastrophic event in the life of a neutron star, and one that initiated the series of bursts subsequently observed. Our observation of the 27 August 1998 event leads to a different interpretation. The source evolved from a weak, infrequent repeater to an intensely active one, indicating that the neutron star's crust was able to adjust to magnetic stresses by undergoing relatively minor, localized cracking for a long period. The small precursor to the giant flare was comparable in intensity to these bursts, and may have been the final trigger for it. In the following months, these bursts have continued. Thus our observations imply that rare giant flares on SGRs may be the rule, rather than the exception, and that they may occur at any time. It therefore seems likely that SGR0525–66 had emitted relatively weak bursts before 5 March 1979, and that these bursts were not detected owing to their weakness or, perhaps, spacecraft coverage. The magnetar theory predicts that on any given SGR, such events may recur on a timescale of decades or more (R. Duncan, personal communication): as it is now almost two decades since the 5 March event, future monitoring of this and other SGRs could confirm this idea. $\square$

Received 23 September; accepted 20 October 1998.

- Hurley, K. in *Proc. 3rd Huntsville Symp.* (eds Kouveliotou, C., Briggs, M. and Fishman, G.) 889–896 (AIP Conf. Proc. 384, Am. Inst. Phys., New York, 1996).
- Mazets, E., Golenetskii, S., Guryan, Yu. & Ilyinskii, V. The 5 March 1979 event and the distinct class of short gamma bursts: are they of the same origin? *Astrophys. Space Sci.* **84**, 173–189 (1982).
- Barat, C. *et al.* Evidence for periodicity in a gamma ray burst. *Astron. Astrophys.* **79**, L24–L25 (1979).
- Braun, R., Goss, W. & Lyne, A. Three fields containing young pulsars: the observable lifetime of supernova remnants. *Astrophys. J.* **340**, 355–361 (1989).
- Thompson, C. & Duncan, R. The soft gamma repeaters as very strongly magnetized neutron stars. I. Radiative mechanism for outbursts. *Mon. Not. R. Astron. Soc.* **275**, 255–300 (1995).
- Mazets, E. *et al.* A flaring X-ray pulsar in Dorado. *Nature* **282**, 587–589 (1979).
- Cline, T. *et al.* Detection of a fast, intense and unusual gamma ray transient. *Astrophys. J.* **237**, L1–L5 (1980).
- Golenetskii, S., Ilyinskii, V., Mazets, E. & Guryan, Yu. Repeating gamma ray bursts from the source FXP 0520–66. *Sov. Astron. Lett.* **5**, 340–342 (1979).
- Golenetskii, S., Ilyinskii, V. & Mazets, E. Recurrent bursts in the source of the 5 March 1979 event. *Nature* **307**, 41–43 (1984).
- Evans, W. *et al.* Location of the gamma-ray transient event of 1979 March 5. *Astrophys. J.* **237**, L7–L9 (1980).
- Cline, T. *et al.* Precise source location of the anomalous 1979 March 5 gamma ray transient. *Astrophys. J.* **255**, L45–L48 (1982).
- Rothschild, R., Kulkarni, S. & Lingenfelter, R. Discovery of an X-ray source coincident with the soft gamma repeater 0525–66. *Nature* **368**, 432–434 (1984).
- Mazets, E., Golenetskii, S. & Guryan, Yu. Soft gamma ray bursts from the source B1900+14. *Sov. Astron. Lett.* **5**(6) 343–344 (1979).
- Kouveliotou, C. *et al.* Recurrent burst activity from the soft gamma-ray repeater SGR 1900+14. *Nature* **362**, 728–730 (1993).
- Kouveliotou, C. *et al.* The rarity of soft gamma-ray repeaters deduced from reactivation of SGR1806–20. *Nature* **368**, 125–127 (1994).
- Vasisht, G., Kulkarni, S., Frail, D. & Greiner, J. Supernova remnant candidates for the soft gamma-ray repeater 1900+14. *Astrophys. J.* **431**, L35–L38 (1994).
- Hurley, K. *et al.* Network synthesis localization of two soft gamma repeaters. *Astrophys. J.* **431**, L31–L34 (1994).
- Hurley, K., Kouveliotou, C., Mazets, E. & Cline, T. *IAU Circ.* No. 6929 (1998).
- Hurley, K. *et al.* Reactivation and precise IPN localization of the soft gamma repeater SGR1900+14. *Astrophys. J.* (in the press).
- Hurley, K., Cline, T., Mazets, E. & Golenetskii, S. *IAU Circ.* No. 7004 (1998).
- Hurley, K., Kouveliotou, C., Murakami, T. & Strohmayer, T. *IAU Circ.* No. 7001 (1998).
- Kouveliotou, C., Strohmayer, T., Hurley, K., van Paradijs, J. & Woods, P. *IAU Circ.* No. 7001 (1998).
- Cline, T., Mazets, E. & Golenetskii, S. *IAU Circ.* No. 7002 (1998).
- Marshall, F. *et al.* *IAU Circ.* No. 7005 (1998).
- Kouveliotou, C. *et al.* An X-ray pulsar with a superstrong magnetic field in the soft gamma-ray repeater SGR1806–20. *Nature* **393**, 235–237 (1998).
- Barat, C. *et al.* Fine time structure in the 1979 March 5 gamma-ray burst. *Astron. Astrophys.* **126**, 400–402 (1983).
- Fenimore, E., Klebesadel, R. & Laros, J. The 1979 March 5 gamma-ray transient: was it a classical gamma-ray burst? *Astrophys. J.* **460**, 964–975 (1996).
- Hurley, K. *et al.* The solar X-ray/cosmic gamma-ray burst experiment aboard Ulysses. *Astron. Astrophys. Suppl. Ser.* **92**, 401–410 (1992).

Acknowledgements. We thank R. Duncan for discussions about the magnetic field estimate. This work was supported in the US by NASA and a contract from JPL, and at the Ioffe Institute by an RSA contract.

Correspondence and requests for materials should be addressed to K.H. (e-mail: khurley@sunspot.ssl.verkeleg.edu).

Reversing the direction of the supercurrent in a controllable Josephson junction

J. J. A. Baselmans*, A. F. Morpurgo*†, B. J. van Wees* & T. M. Klapwijk*

* Department of Applied Physics and Material Science Center, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

When two superconductors are connected by a weak link, a supercurrent flows, the magnitude of which is determined by the difference in the macroscopic quantum phases of the superconductors. This phenomenon was discovered by Josephson¹ for the case of a weak link formed by a thin tunnel barrier: the supercurrent, I , is related to the phase difference, ϕ , through the Josephson current–phase relation, $I = I_c \sin \phi$, with I_c being the critical current which depends on the properties of the weak link. A similar relation holds for weak links consisting of a normal metal, a semiconductor or a constriction². In all cases, the phase difference is zero when no supercurrent flows through the junction, and increases monotonically with increasing supercurrent

until the critical current is reached. Here we use nanolithography techniques to fabricate a Josephson junction with a normal-metal weak link in which we have direct access to the microscopic current-carrying electronic states inside the link. We find that the fundamental Josephson relation can be changed from $I = I_c \sin \phi$ to $I = I_c \sin(\phi + \pi)$ —that is, a π -junction—by controlling the energy distribution of the current-carrying states in the normal metal. This fundamental change in the way these Josephson junctions behave has potential implications for their use in superconducting electronics as well as in (quantum) logic circuits based on superconductors.

The microscopic mechanism responsible for the supercurrent in a Josephson junction is the transport of correlated electrons. In a superconductor/normal-metal/superconductor (SNS) junction, conduction electrons mediate current transport from superconductor 1 (S1) to superconductor 2 (S2) by either ballistic or diffusive transport through the normal metal (N). In a ballistic junction, in which the elastic mean free path is larger than the length of the normal region, Andreev bound states are formed^{3–5}. The dispersion relation of these states is such that each subsequent state carries a supercurrent in the positive or negative direction at a given value of the macroscopic phase difference between the superconductors; the states are degenerate if the phase is zero. The net supercurrent that flows between the two superconductors depends therefore not only on the actual phase difference ϕ , but also on the occupation of the Andreev bound states. The prediction is that the electron energy distribution function in the normal region will change the supercurrent, even resulting in a sign reversal^{6–8}.

Transport of electrons in metals is usually diffusive, the electron trajectories are not well defined, and Andreev bound states are no longer the natural concept to describe the supercurrent. But electron correlations induced by the superconducting electrodes are still present, with the energy scale determined by the Thouless energy $E_T = \hbar D/L^2$, where D is the diffusion coefficient and L is the separation between the superconductors. The energy spectrum of the superconducting correlations is expressed in a so-called supercurrent-carrying density of states, which can be calculated directly using the quasiclassical Green's function theory of superconductivity^{9–12}. The supercurrent-carrying density of states is an odd function of energy; it shows a phase-dependent mini-gap at low energies, above which it has a positive maximum, after which it changes sign and approaches zero at high energies. The positive and negative

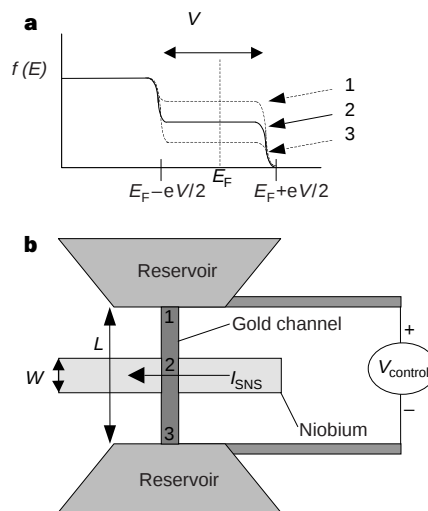


Figure 1 Electronic distribution function and the sample layout. In the bottom panel, a gold channel between two electron reservoirs is connected to two niobium superconducting leads. The control voltage across the channel induces a position-dependent electron distribution, shown in **a** for positions 1, 2 and 3 in **b**. The current through the Josephson junction is indicated by I_{SNS} .

† Present address: Department of Physics, Stanford University, Stanford, California 94305-4060, USA.

parts of the supercurrent-carrying density of states represent, at a given phase, energy-dependent contributions to the supercurrent in the positive and negative direction. The size and direction of the total supercurrent depends therefore on the occupied fraction of these states, which is analogous to the occupation of the discrete Andreev bound states in a ballistic system. In a recent experiment, Morpurgo *et al.*¹³ changed the occupied fraction of the supercurrent-carrying density of states by raising the effective electron temperature. This was done by sending a normal current from normal reservoirs through the N-part of the SNS junction, leading to a Fermi–Dirac distribution with the electrons at a slightly elevated temperature. A monotonic decrease of the supercurrent was observed, which is expected for a thermal distribution.

In a mesoscopic wire, however, the distribution of electrons over the energies is not necessarily thermal. As shown by Pothier *et al.*^{14,15}, the electron distribution in a normal wire attached to two large electron reservoirs at a voltage difference V may have a double-step structure. A non-equilibrium state is reached by changing the chemical potentials of the electron reservoirs in opposite directions by applying a control voltage. The energy of the electrons in the wire depends on the distribution functions of the reservoirs, possibly modified by inelastic relaxation processes inside the wire. If the wire is sufficiently short, so that the diffusion time τ_D is smaller than the inelastic scattering time τ_i , the energy of the electrons will be conserved over the length of the wire. The energy distribution of the electrons in the channel is then given by the superposition of the Fermi–Dirac distributions of the reservoirs. This results, at sufficiently low temperatures $k_B T \ll eV$ (where k_B is Boltzmann's constant and e the single electron charge), in a position-dependent double-step function (Fig. 1a). The energy separation between the steps is eV . As predicted^{9–12}, such an energy distribution will have a profound effect on the supercurrent in an SNS junction, even reversing its direction for sufficiently large control voltages. The microscopic mechanism responsible for this is the redistribution of the occupied fraction of the current-carrying density of states due to the change in the electronic distribution function inside the junction. The stable zero-current state, which corresponds to $\phi = 0$ for a conventional Josephson junction, corresponds to $\phi = \pi$ at large control voltages, resulting in a current–phase relation given by $I = I_c \sin(\phi + \pi)$, hence the name π -junction. This mechanism should not be confused with the π -junction behaviour induced by

magnetic impurities (refs 16, 17 and references therein) or resulting from the symmetry of the order parameter in ceramic superconductors (ref. 18 and references therein). Not only are the microscopic mechanisms different in these cases, but in our case the junction is also controllable, its state depending on the applied control voltage.

We study the behaviour of the Josephson current in four similar diffusive SNS junctions by applying a non-thermal distribution function to the normal conducting weak link. We present the results of one device. The device is shown in Fig. 1b. A thin (~ 40 nm) and narrow (~ 200 nm) gold channel is connected to two relatively thick (475 nm) electron reservoirs of millimetre size. In the other direction, the gold channel is coupled in its centre to two superconducting niobium electrodes. The distance between the two niobium electrodes is 300 nm. The measurements are performed at 100 mK to realize a sharp Fermi–Dirac distribution function inside the reservoirs. Electronic filtering at room temperature and at 100 mK is used to reduce external noise and hence, unintentional heating of the electrons. The gold of the channel has a diffusion coefficient at $180 \text{ cm}^2 \text{ s}^{-1}$, resulting in a Thouless energy $E_T \approx 140 \mu\text{eV}$, so that at the measuring temperature $k_B T = 10 \mu\text{eV} < E_T < \Delta = 1,500 \mu\text{eV}$ (here Δ is the superconducting energy gap of niobium). The length of the control channel between the reservoirs, L , is $1 \mu\text{m}$, resulting in an estimated diffusion time of $\tau_D = 50$ ps. This is much smaller than the estimated energy relaxation time for electrons in a thin film of impure gold¹⁹. The reservoirs are of millimetre size because they also need to act as effective cooling fins to prevent unwanted electron heating and hence maintain the desired step-like electron distribution²⁰. The contacts between the gold and the niobium are cleaned using argon sputter etching to ensure a high interface transparency.

In the experiment, the current–voltage (I – V) curves of the SNS junction are measured for different values of control voltage across the control channel (V_{control}). From these curves, we determine the critical current of the junction as a function of the control voltage (Fig. 2). The critical current of the SNS junction shows a non-monotonic behaviour as a function of control voltage. At zero control voltage, the product $I_c R_n$ (where R_n is the normal state resistance of the junction), $200 \mu\text{V}$, is of the order of the Thouless energy, in good agreement with the theory of diffusive SNS junctions²¹. When a voltage is applied, the distribution function

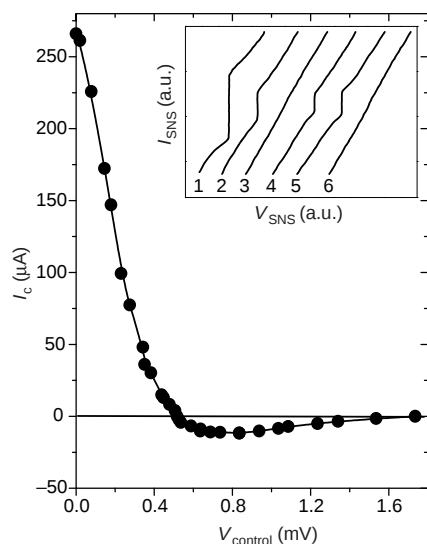


Figure 2 Critical current of the SNS junction as a function of the control voltage. The current is plotted negative above $V_{\text{critical}} = 0.52$ mV because of the π -state of the junction; a.u., arbitrary units. Inset, selected I – V curves for the following control voltages: 0.38 mV (curve 1), 0.44 mV (2), 0.52 mV (3), 0.64 mV (4), 0.84 mV (5), 1.70 mV (6).

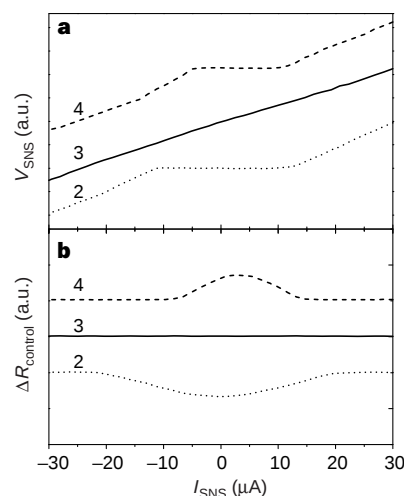


Figure 3 V_{SNS} , the voltage across the junction, and the resistance of the control channel as a function of the current I_{SNS} for three different values of the control voltage. The curves in **a** and **b** correspond to curves 2, 3 and 4 of Fig. 2 inset (offset for clarity). The resistance variation changes sign, although the corresponding I – V curves look similar for $V_{\text{control}} < V_{\text{critical}}$ (2) or $V_{\text{control}} < V_{\text{critical}}$ (4).

inside the gold wire changes, and therefore the occupation of the energy levels carrying the supercurrent also changes. A decrease of the supercurrent is observed, reaching zero at $V_{\text{control}} = 520 \mu\text{V} \equiv V_{\text{critical}}$, corresponding to approximately four times the Thouless energy; no sign of a supercurrent can be detected (Fig. 2, inset). At higher voltages, the supercurrent reappears. In Fig. 2 we plot the critical supercurrent in this region with a minus sign, anticipating that the junction has entered the π -state. Above voltages of 1.7 meV, no supercurrent can be detected.

To determine the current–phase relation, $I = I_c \sin(\phi)$ or $I = I_c \sin(\phi + \pi)$, we need an independent way to determine the phase difference across the junction for a fixed direction of the applied supercurrent. We recall that the normal conductance of the control channel is also modulated by the phase difference ϕ —a phenomenon known as Andreev interferometry and extensively studied in recent years^{22,23}. The modulation of the conductance arises from the fact that Andreev reflected holes scatter from both superconductors with a different phase shift, and therefore might interfere constructively ($\phi = 0$) or destructively ($\phi = \pi$). This leads to a phase-dependent change in the diffusion constant in N. The resulting phase dependence of the resistance is given approximately by $\Delta R = -R_0(1 + \cos(\phi))$, where R_0 has a temperature- and energy-dependent amplitude.

In Fig. 3 we show the resistance of the normal control channel (Fig. 3b) and the observed I – V curves of the SNS junction (Fig. 3a). The dependence of ΔR on the applied current changes sign on crossing the critical control voltage. The middle curve (3) shows no modulation at all; the corresponding I – V curve is linear. If we suppose that the current–phase relation is given by the conventional Josephson relation $I = I_c \sin(\phi)$, then the phase difference between the two superconductors changes from $-\pi/2$ through 0 to $\pi/2$ if the current through the junction I_{SNS} is varied from $-I_c$ to $+I_c$. Hence the resistance of the control channel will show a minimum at $I_{\text{SNS}} = 0$, corresponding to $\phi = 0$. This is observed for all control voltages $V_{\text{control}} < V_{\text{critical}}$. The upper curve in Fig. 3b is an example of the behaviour for $V_{\text{control}} > V_{\text{critical}}$: a maximum is observed in the resistance for $I_{\text{SNS}} = 0$, consistent with the assumption that the current–phase relation is now given by $I = I_c \sin(\phi + \pi)$ and $\phi = \pi$ when $I_{\text{SNS}} = 0$. The phase difference π between the two superconductors appears whenever $V_{\text{control}} > V_{\text{critical}}$. We take this as direct proof that the junction switches from a normal state to a π -state as a function of the control voltage.

The qualitative dependence of I_c on V_{control} and the energy at which the phase jumps, are both in agreement with theoretical work on diffusive SNS junctions^{10–12}. However, the relative magnitude of the supercurrent in the π -state range of V_{control} is smaller than one would expect in comparison to I_c at $V_{\text{control}} = 0$. The most important reason for this is probably the geometry of the sample. In our case, the width-to-length ratio W/L is ~ 0.4 , so that the distribution function is not constant over the entire junction length. This will reduce the magnitude of the supercurrent in the π -regime, as predicted theoretically⁹. □

Received 9 July; accepted 30 September 1998.

- Josephson, B. D. Possible new effects in superconducting tunneling. *Phys. Lett.* **1**, 251–253 (1962).
- Likharev, K. K. Superconducting weak links. *Rev. Mod. Phys.* **51**, 101–169 (1979).
- Kulik, I. O. Macroscopic quantization and the proximity effect in S–N–S junctions. *Zh. Eksp. Teor. Fiz.* **57**, 1745–1750 (1969); *Sov. Phys. JETP* **30**, 944–950 (1970).
- Ishii, C. Josephson currents through junctions with normal metal barriers. *Prog. Theor. Phys.* **5**, 1525–1547 (1972).
- Bardeen, J. & Johnson, J. L. Josephson current flow in pure superconducting–normal–superconducting junctions. *Phys. Rev. B* **5**, 72–78 (1972).
- Van Wees, B. J., Lennes, K.-M. H. & Harmans, C. J. P. M. Transmission formalism for supercurrent flow in multiprobe superconductor–semiconductor devices. *Phys. Rev. B* **44**, 470–473 (1991).
- Wendin, G. & Shumeiko, V. S. Giant Josephson current through a single bound state in a superconducting tunnel junction. *Phys. Rev. B* **53**, R6006–R6009 (1996).
- Chang, L.-F. & Bagwell, P. F. Control of Andreev level occupation in a Josephson junction by a normal metal probe. *Phys. Rev. B* **55**, 12678–12690 (1997).
- Volkov, A. F. New phenomena in Josephson SINIS junctions. *Phys. Rev. Lett.* **74**, 4730–4733 (1995).
- Volkov, A. F. & Takayanagi, H. Long range phase coherent effects in the transport properties of mesoscopic superconductor–normal-metal structures. *Phys. Rev. B* **56**, 11184–11194 (1997).
- Wilhelm, F. K., Schön, G. & Zaikin, A. D. Mesoscopic superconducting–normal metal–superconducting transistor. *Phys. Rev. Lett.* **81**, 1682–1685 (1998).

- Yip, S.-K. Energy resolved supercurrent between two superconductors. *Phys. Rev. B* **58**, 5803–5807 (1998).
- Morpurgo, A. E., Klapwijk, T. M. & van Wees, B. J. Hot electron tunable supercurrent. *Appl. Phys. Lett.* **72**, 966–968 (1998).
- Pothier, H., Guéron, S., Birge, N. O., Esteve, D. & Devoret, M. H. Energy distribution function of quasiparticles in mesoscopic wires. *Phys. Rev. Lett.* **79**, 3490–3493 (1997).
- Pothier, H., Guéron, S., Birge, N. O., Esteve, D. & Devoret, M. H. Energy distribution of electrons in an out-of-equilibrium metallic wire. *Z. Phys. B* **104**, 178–182 (1997).
- Bulaevski, L. N., Kuzii, V. V. & Sobyenin, A. A. On possibility of spontaneous magnetic flux in a Josephson junction containing magnetic impurities. *Solid State Commun.* **25**, 1053–1057 (1978).
- Mühge, Th. *et al.* Possible origin for oscillatory superconducting transition temperature in superconducting/ferromagnet multilayers. *Phys. Rev. Lett.* **77**, 1857–1860 (1996).
- van Harlingen, D. J. Phase-sensitive tests of the symmetry of the pairing state in high-temperature superconductors—evidence for $d_{x^2-y^2}$ symmetry. *Rev. Mod. Phys.* **67**, 515–535 (1995).
- Echternach, P. M., Ghershenson, M. E. & Bozler, M. H. Evidence of interference between electron-phonon and electron-impurity scattering on the conductivity of thin metal films. *Phys. Rev. B* **47**, 13659–13663 (1993).
- Wellstood, E. C., Urbina, C. & Clarke, J. Hot electron effects in metals. *Phys. Rev. B* **49**, 5942–5955 (1994).
- Wilhelm, F. K., Zaikin, A. D. & Schön, G. Supercurrent in a mesoscopic proximity wire. *J. Low Temp. Phys.* **106**, 305–310 (1997).
- Esteve, D. *et al.* in *Mesoscopic Electron Transport: NATO ASI series E*, vol. 345 375–406 (eds Sohn, L. L. *et al.*) (Kluwer Academic, Dordrecht/Boston/London, 1997).
- Van Wees, B. J. & Takayanagi, H. in *Mesoscopic Electron Transport: NATO ASI series E*, vol. 345, 469–502 (eds Sohn, L. L. *et al.*) (Kluwer Academic, Dordrecht/Boston/London, 1997).

Acknowledgements. We thank F. K. Wilhelm for discussions. This work was supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) through the Stichting voor Fundamenteel Onderzoek der Materie (FOM).

Correspondence and requests for materials should be addressed to J.J.A.B. (e-mail: Baselman@phys.rug.nl).

Scaling of transition temperature and CuO_2 plane buckling in a high-temperature superconductor

O. Chmaissem^{††}, J. D. Jorgensen[†], S. Short[†], A. Knizhnik[‡], Y. Eckstein[‡] & H. Shaked[§]

^{*} Science and Technology Center for Superconductivity, and [†] Materials Science Division, Argonne National Laboratory, Argonne, Illinois 60439, USA

[‡] Department of Physics and Crown Center for Superconductivity, Technion, Haifa 32000, Israel

[§] Department of Physics, Ben Gurion University, Beer-Sheva 84105, Israel

A characteristic feature of the high-temperature superconductors is the existence of a chemical composition that gives a maximum transition temperature, T_c , separating the so-called under-doped and over-doped regimes^{1,2}. This behaviour is thought to be universal for high-temperature superconductors. In practice, there are only a few high- T_c compounds for which the composition can be varied continuously throughout the entire doping range. Here we report a study of correlations between structure and T_c in a compound with the ‘123’ structure in which both the under-doped and over-doped regimes can be accessed. We observe a clear scaling between T_c and the buckling of the copper oxide planes; both go through a maximum at the same oxygen composition (and hence doping level), so implying a common origin. Previous work has shown that, for a fixed chemical composition, increased CuO_2 plane buckling lowers the transition temperature^{3–11}. Thus the observation of a maximum in the buckling at the maximum T_c indicates that, as the composition is changed to increase T_c , there is a structural response that competes with superconductivity.

The 123 compound used for these studies is $(\text{La}_{1-x}\text{Ca}_x)(\text{Ba}_{1.75-x}\text{La}_{0.25+x})\text{Cu}_3\text{O}_y$ (ref. 12). In previous work¹², it has been shown that this compound can be readily made as a single phase, with the cation substitutions as indicated by the formula, for $0 \leq x \leq 0.4$. Powder samples of $(\text{La}_{1-x}\text{Ca}_x)(\text{Ba}_{1.75-x}\text{La}_{0.25+x})\text{Cu}_3\text{O}_y$ with $x = 0.1$ and 0.4 were synthesized by the methods described previously¹². The

samples were confirmed to be single phase, within the limits of detectability, by X-ray and neutron diffraction. Oxygen contents were varied by subsequent annealing of parts of these samples in oxygen pressures from 10^{-4} to 1,100 atm at 300–400 °C. Values of T_c were obtained from measurements of magnetic susceptibility (Fig. 1) and resistivity.

The structures of all samples were determined by Rietveld refinement of neutron powder diffraction data from the Special Environment Powder Diffractometer at Argonne's Intense Pulsed Neutron Source; the same structural model¹³ (tetragonal, space

group $P4/mmm$, at all oxygen compositions) was used that had previously been applied to 123 compounds with similar cation substitutions. A typical Rietveld refinement profile is shown in Fig. 2. The refinements confirm the cation substitutions consistent with the starting compositions, and give refined values of the total oxygen contents that agree with values determined by iodometric titration.

The tetragonal structure for all oxygen compositions is typical of a 123 compound with significant partial substitution of a trivalent cation at the Ba site^{13,14}. The La^{3+} ion at the Ba site bonds oxygen atoms into both available oxygen sites in the Cu1 plane (that is, the sites that would normally be the chain and anti-chain sites); this disrupts the formation of long-range ordered Cu–O chains and results in an average tetragonal structure^{13,14} in which the two sites are symmetry equivalent. Thus the Cu1 plane has a single oxygen site that can accommodate up to two oxygen atoms per formula unit. The total oxygen content would be eight atoms per formula unit if this site were filled. A second feature of such cation substitutions is that the oxygen solubility limit is raised above seven atoms per formula unit. $(\text{La}_{1-x}\text{Ca}_x)(\text{Ba}_{1.75-x}\text{La}_{0.25+x})\text{Cu}_3\text{O}_y$ remains stable under oxygen annealing at elevated pressures, and an oxygen content corresponding to $y = 7.3$ can be achieved by annealing at 300 °C in oxygen at 1,100 atm.

The structural results confirm that the charge transfer to the CuO_2 planes increases smoothly from the under-doped regime, through the maximum T_c , into the over-doped regime. The lattice parameters vary monotonically with oxygen content (Fig. 3a). The Cu–O apical bond length, which has been shown to be a measure of the charge transfer¹⁵, also varies monotonically, as does the hole concentration for the CuO_2 planes determined from bond valence sums¹⁶ (Fig. 3b).

A well-defined maximum in T_c is seen at an oxygen content around $y = 7.15$ (Fig. 4). The buckling angle of the CuO_2 planes goes through a maximum at the same oxygen content, implying that the behaviours of buckling and T_c versus doping have a common origin. The observation of maximum CuO_2 buckling at the maximum T_c at first appears to contradict the previous conclusion about the relationship between plane buckling and T_c . In the 123 system and other high- T_c superconductors, there are a number of results that show that, for a constant doping level, increased CuO_2 plane buckling decreases T_c (refs 3–11). The apparent contradiction

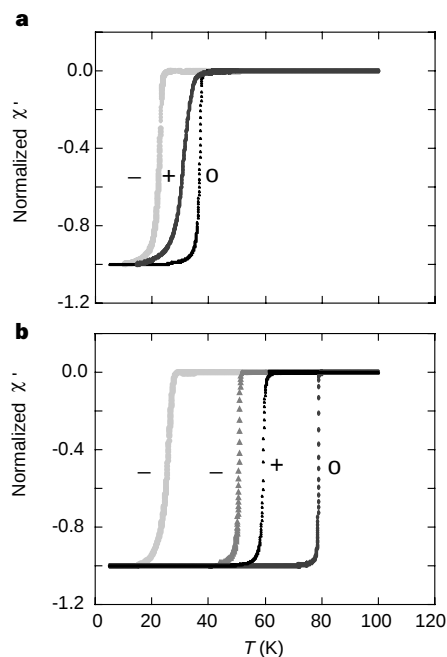


Figure 1 Magnetic susceptibility data for representative under-doped (–), optimally doped (o), and over-doped (+) samples. Compositions shown were $(\text{La}_{1-x}\text{Ca}_x)(\text{Ba}_{1.75-x}\text{La}_{0.25+x})\text{Cu}_3\text{O}_y$, with $x = 0.1$ (a) and $x = 0.4$ (b). χ' is the real part of the magnetic susceptibility. The measured values have been normalized such that $\chi' = -1$ for the saturated value in the superconducting state. Onset T_c s are defined as the linear extrapolation from the midpoint of the transition to $\chi' = 0$.

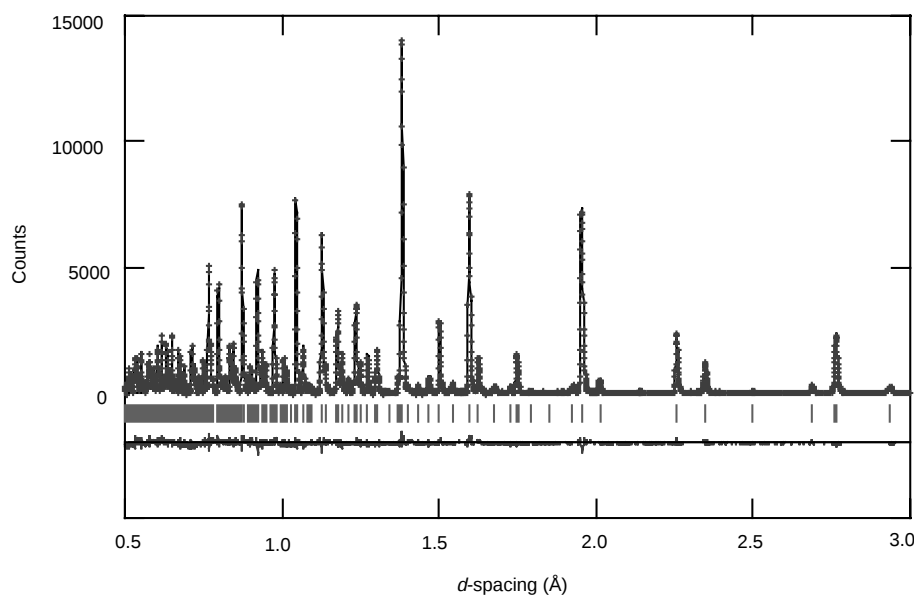


Figure 2 Best-fit Rietveld refinement profile showing observed (+ symbols) and calculated (line) intensities. The markers below the profile correspond to Bragg peak positions for $(\text{La}_{1-x}\text{Ca}_x)(\text{Ba}_{1.75-x}\text{La}_{0.25+x})\text{Cu}_3\text{O}_y$ with $x = 0.1$ and

$y = 7.08 (\pm 0.02)$. The difference between observed and calculated values is shown at the bottom. The quality of the fit confirms that the sample is single phase and the structural model used for refinement is appropriate.

between our result and previous work arises from the different experimental strategies. The previous conclusion is based on experiments where the relationship between T_c and buckling was studied at a constant chemical composition (constant doping) and on the observation that the highest maximum T_c s are achieved in compounds with the smallest buckling. We have probed the behaviour of buckling and T_c when both are responses to a varying doping level. Combining our result with the previously published work, we conclude that, as the composition is varied to increase the T_c , there is a structural response that competes with superconductivity, that is, there is a structural distortion (buckling of the CuO_2 planes) of a nature that is known to decrease T_c .

This competition between superconductivity and structural distortion is reminiscent of behaviour observed in many conventional superconductors. However, in those systems, the structural distortion often results in a phase transition that destroys superconductivity. For example, in the Chevrel-phase superconductors, MMo_6S_8 , as chemical substitutions are used to increase T_c by changing the size of the M ion, there is a symmetry-lowering transition, which opens a gap at the Fermi energy and destroys metallic behaviour^{17–19}. Similarly, in the cubic $\text{Ba}_{1-x}\text{K}_x\text{BiO}_3$ system²⁰, as the K content is decreased to increase T_c , there is a transition to a non-superconducting phase with a lower-symmetry structure. The explanation proposed for this behaviour in conventional superconductors was that the structure distorted in order to lower the free energy of the system as one attempted to increase T_c by approaching a peak in the electronic density of states. In contrast to the conventional superconductors, in the 123 system studied here, there is a structural distortion, known to lower T_c , that grows in magnitude as the composition is varied to approach the maximum T_c . However, the effect of this distortion is sufficiently weak that the maximum in T_c remains well defined and a transition to a non-superconducting state never occurs.

The present experiments do not identify the cause of the coupled

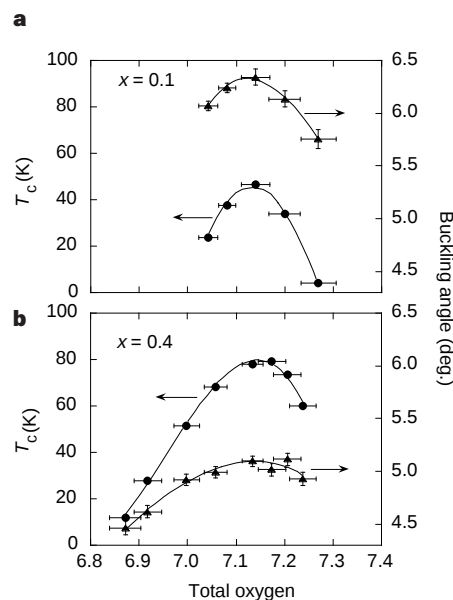


Figure 4 Variation of T_c and buckling angle of the CuO_2 plane with total oxygen content. The buckling angle is defined as the angle by which the plane oxygen atoms are out of the plane of the copper atoms. Compositions shown are $(\text{La}_{1-x}\text{Ca}_x)(\text{Ba}_{1.75-x}\text{La}_{0.25+x})\text{Cu}_3\text{O}_y$ with $x = 0.1$ (**a**) and $x = 0.4$ (**b**). This scaling of the CuO_2 plane buckling angle with T_c implies that these two responses to the varying doping have a common origin. We also note that the highest maximum T_c occurs for the series with the smallest maximum buckling angle ($x = 0.4$), which is consistent with the previous conclusion^{3–11} that plane buckling decreases T_c for a constant doping. When combined, these two observations lead to the conclusion that the buckling is a response to the varying doping that competes with superconductivity. Oxygen contents are determined from Rietveld refinement of neutron powder diffraction data.

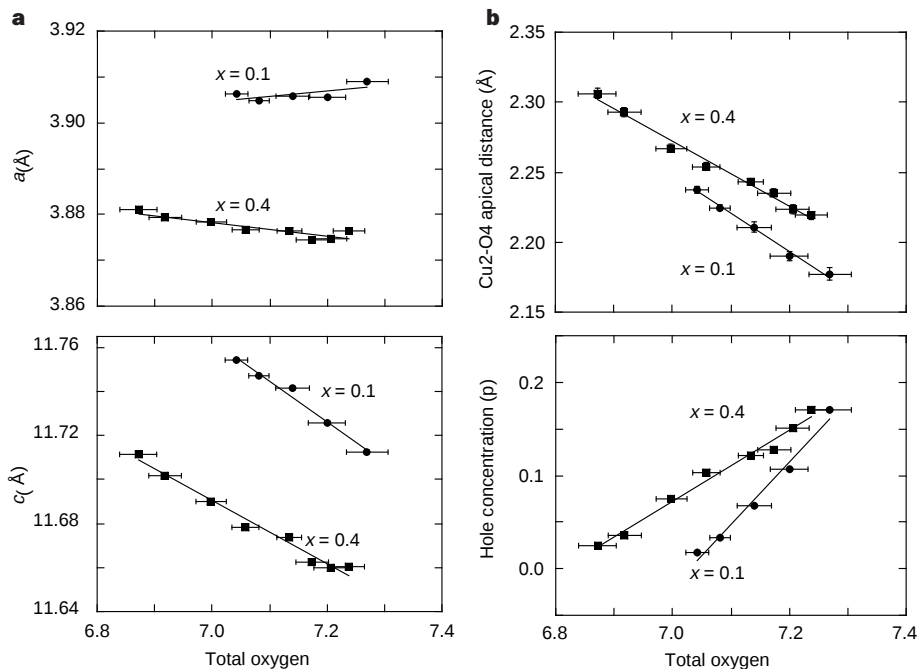


Figure 3 Variation of lattice parameters and apical copper-oxygen bond length with total oxygen content. Shown are lattice parameters (**a**), and Cu2-O4 apical copper-oxygen bond length and hole concentration per CuO_2 plane (**b**) as determined by bond valence sum calculations using the bond lengths obtained from Rietveld refinements. Compositions shown are

$(\text{La}_{1-x}\text{Ca}_x)(\text{Ba}_{1.75-x}\text{La}_{0.25+x})\text{Cu}_3\text{O}_y$ with $x = 0.1$ (filled circles) and $x = 0.4$ (filled squares). The smooth variation of these parameters with oxygen content confirms that the charge transfer to the CuO_2 planes increases continuously across the entire composition range. Oxygen contents are determined from Rietveld refinements.

variations in T_c and CuO_2 plane buckling. However, a consistent explanation can be proposed on the basis of what is generally known about the electronic structures of high- T_c compounds. Shortly after the discovery of high- T_c superconductivity in the layered copper oxides, it was proposed that the maximum T_c corresponds to a peak in the electronic density of states associated with flat regions of the Fermi surface^{21–24}. Electronic band structure calculations showed that the actual structures of the high- T_c compounds—specifically, the CuO_2 planes, and, in the case of the $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ compound, the CuO chains—resulted in the extension of these flat regions of the Fermi surface, placing more states in the peak²⁵. This was confirmed experimentally, for example, for the $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ compound, by high-resolution angle-resolved photoemission studies^{26,27}.

In such a picture, the maximum of the CuO_2 plane buckling at the maximum T_c is a response to the electronic structure at this composition. Intuitively, the buckling of the CuO_2 planes is the structural parameter most sensitive to the relevant features of the electronic structure. Band structure calculations²⁵ and photoemission studies^{26,27} for the $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ structure show that the CuO_2 planes and the one-dimensional CuO chains both contribute to the peak in the density of states, but that the chain contribution diminishes as the chain oxygen is depleted and disordered²⁸. The 123 compound we have studied is tetragonal and does not contain ordered CuO chains. At most, it can have only disordered short chain fragments (extending only a few unit cells for the La/Ba compositions studied here), consistent with the average tetragonal symmetry. Its electronic structure has not been calculated, but we assume that the CuO_2 planes are the dominant contribution to the peak in the density of states²⁹. Buckling would be expected to modify the Fermi surface so as to lower and/or broaden the peak, as has been calculated³⁰ and shown experimentally^{3–7} for the La_2CuO_4 structure. Thus the observed changes in buckling angle versus oxygen content indicate that the structure distorts to lower its free energy as the Fermi energy passes through the peak in the density of states. If this hypothesis is correct, the present experiments use a structural feature—the buckling of the CuO_2 planes—to provide evidence that the maximum T_c is located at a peak in the electronic density of states. □

Received 9 July; accepted 14 October 1998.

- Whangbo, M.-H., Kang, D. B. & Torardi, C. C. Correlations between the superconducting transition temperatures and the in-plane copper-oxygen bond lengths in the copper-oxide superconductors. *Physica C* **158**, 371–376 (1989).
- Zhang, H. & Sato, H. Universal relationship between T_c and the hole content in p-type cuprate superconductors. *Phys. Rev. Lett.* **70**, 1697–1699 (1993).
- Büchner, B., Breuer, M., Freimuth, A. & Kampf, A. P. Critical buckling for the disappearance of superconductivity in rare-earth-doped $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$. *Phys. Rev. Lett.* **73**, 1841–1844 (1994).
- Yamada, Y. & Ido, M. Pressure effects on superconductivity and structural phase transitions in $\text{La}_{2-x}\text{M}_x\text{CuO}_4$ ($\text{M}=\text{Ba}, \text{Sr}$). *Physica C* **203**, 240–246 (1992).
- Takahashi, H. *et al.* Structural effects of hydrostatic pressure in orthorhombic $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$. *Phys. Rev. B* **50**, 3221–3229 (1994).
- Dabrowski, B. *et al.* Dependence of superconducting transition temperature on doping and structural distortion of the CuO_2 planes in $\text{La}_{2-x}\text{M}_x\text{CuO}_4$ ($\text{M}=\text{Nd}, \text{Cr}, \text{Sr}$). *Phys. Rev. Lett.* **76**, 1348–1351 (1996).
- Braden, M. *et al.* Coupling between superconductivity and structural deformation in $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ ($x \sim 0.13$). *Phys. Rev. B* **47**, 12288–12291 (1993).
- Kaldis, E. *et al.* A displacive structural transformation in the CuO_2 planes of $\text{YBa}_2\text{Cu}_3\text{O}_x$ at the underdoped-overdoped phase separation line. *Phys. Rev. Lett.* **79**, 4894–4897 (1997).
- Krüger, Ch., Conder, K., Schwer, H. & Kaldis, E. The dependence of the lattice parameters on oxygen content in orthorhombic $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$: A high precision reinvestigation of near equilibrium samples. *J. Solid State Chem.* **134**, 356–361 (1997).
- Radaelli, P. G. *et al.* Structure, doping and superconductivity in $\text{HgBa}_2\text{CaCu}_2\text{O}_{6+\delta}$ ($T_c \leq 128\text{K}$). *Physica C* **216**, 29–35 (1993).
- Jorgensen, J. D. *et al.* in *Recent Developments in High Temperature Superconductivity* (eds Klamut, J., Veal, B. W., Dabrowski, B. M. & Klamut, P. W.) 1–16 (Lecture Notes in Physics, Springer, Berlin, 1996).
- Goldschmidt, D., Knizhnik, A., Direktovitch, Y., Reisner, G. M. & Eckstein, Y. Variation of T_c and resistivity in charge-compensated $(\text{Ca}, \text{La}_{1-x})\text{Ba}_{1+x}\text{Cu}_2\text{O}_7$. *Phys. Rev. B* **49**, 15928–15935 (1994).
- Dabrowski, B. *et al.* New 1212 structural analog of $\text{YBa}_2\text{Cu}_3\text{O}_7$. The Sr-based superconducting copper oxide $\text{CuSr}_{2-x}\text{La}_x\text{YCu}_2\text{O}_{7+\delta}$ with $T_c \sim 60\text{K}$. *Physica C* **208**, 183–188 (1993).
- Kramer, M. J. *et al.* Hole filling charge transfer and superconductivity in $\text{Nd}_{1+x}\text{Ba}_{2-x}\text{Cu}_3\text{O}_{7+\delta}$. *Physica C* **219**, 145–155 (1994).
- Jorgensen, J. D. *et al.* Structural properties of oxygen-deficient $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$. *Phys. Rev. B* **41**, 1863–1877 (1990).
- Tallon, J. L. The relationship between bond-valence sums and T_c in cuprate superconductors. *Physica C* **168**, 85–90 (1990).

- Baillif, R., Dunand, A., Müller, J. & Yvon, K. Structural phase transformation in the cluster chalcogenides EuMo_2S_8 and BaMo_2S_8 . *Phys. Rev. Lett.* **47**, 672–675 (1981).
- Lachal, B., Baillif, R., Junod, A. & Müller, J. Structural instabilities of chevre phases: The alkaline earth molybdenum sulphide series. *J. Solid State Commun.* **45**, 849–851 (1983).
- Jorgensen, J. D., Hinks, D. G. & Felcher, G. P. Lattice instability and superconductivity in the Pb, Sn, and Ba Chevrel phases. *Phys. Rev. B* **35**, 5365–5368 (1987).
- Pei, S. *et al.* Structural phase diagram of the $\text{Ba}_{1-x}\text{K}_x\text{BiO}_3$ system. *Phys. Rev. B* **41**, 4126–4141 (1990).
- Hirsh, J. & Scalapino, D. J. Enhanced superconductivity in quasi two-dimensional systems. *Phys. Rev. Lett.* **56**, 2732–2735 (1986).
- Labbé, J. & Bok, J. Superconductivity in alkaline-earth-substituted La_2CuO_4 : A theoretical model. *Europhys. Lett.* **3**, 1225–1230 (1987).
- Xu, J.-H., Watson-Yang, T. J., Yu, J. & Freeman, A. J. Dominant role of the 2D Van Hove singularity on the Fermi surface and generalized susceptibility of the quasi-2D superconductor $\text{La}_{2-x}\text{M}_x\text{CuO}_4$ ($\text{M}=\text{Sr}, \text{Ba}, \dots$). *Phys. Lett. A* **120**, 489–493 (1987).
- Friedel, J. On the broadening of low dimension Van Hove singularities in oxide superconductors. *J. Phys. (Paris)* **49**, 1435–1441 (1988).
- Pickett, W. E. Electronic structure of the high-temperature oxide superconductors. *Rev. Mod. Phys.* **61**, 433–512 (1989).
- Liu, R. *et al.* Electronic structure near E_F in $\text{YBa}_2\text{Cu}_3\text{O}_x$ for $6.35 \leq x \leq 6.9$: A photoemission study. *Phys. Rev. B* **45**, 5614–5621 (1992).
- Liu, R. *et al.* Fermi-surface topology of $\text{YBa}_2\text{Cu}_3\text{O}_x$ with varied oxygen stoichiometry: A photoemission study. *Phys. Rev. B* **46**, 11056–11068 (1992).
- Zaenen, J., Paxton, A. T., Jepsen, O. & Andersen, O. K. Chain-fragment doping and the phase-diagram of $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$. *Phys. Rev. Lett.* **60**, 2785–2688 (1988).
- Szotek, Z., Gyorffy, B. L., Temmerman, W. M. & Andersen, O. K. Van Hove scenario and the eight-band model for high- T_c superconductors. *Phys. Rev. B* **58**, 522–526 (1998).
- Norman, M. R., McMullan, G. J., Novikov, D. L. & Freeman, A. J. Effect of structure on the electronic density of states of doped lanthanum cuprate. *Phys. Rev. B* **48**, 9935–9937 (1993).

Acknowledgements. This work was supported by the National Science Foundation, Office of Science and Technology Centers; the US Department of Energy, Division of Basic Energy Sciences-Materials Sciences; by the Israel Science Foundation administered by Israel Academy of Sciences and Humanities and by the Center of Absorption in Science, Ministry of Immigrant Absorption, Israel.

Correspondence and requests for materials should be addressed to J.D.J. (e-mail: JJorgensen@anl.gov).

Molecular growth from a Mo_{176} to a Mo_{248} cluster

A. Müller, Syed Q. N. Shah, H. Bögge & M. Schmidtman

Lehrstuhl für Anorganische Chemie I. Fakultät für Chemie, Universität Bielefeld, Postfach 100131, D-33501 Bielefeld, Germany

In polyoxometalate chemistry a large variety of compounds, clusters and solid-state structures can be formed by the linking together of well-defined metal–oxygen building blocks^{1,2}. These species exhibit unusual topological and electronic properties, and find applications ranging from medicine³ to industrial processes⁴. The recently reported ring-shaped mixed-valence polyoxomolybdates of the type $\{\text{Mo}_{154}\}$ (refs 5, 6) and $\{\text{Mo}_{176}\}$ (refs 7, 8) represent a new class of giant clusters with nanometre-sized cavities and interesting properties for host–guest chemistry. Here we describe the formation of related clusters of the type $\{\text{Mo}_{248}\}$ formed by addition of further units to the inner surface of the $\{\text{Mo}_{176}\}$ ‘wheel’. The additional units arrange themselves into two $\{\text{Mo}_{36}\}$ ‘hub-caps’ on the initial wheel—clusters that are not stable in isolation. These findings reveal a new pathway to the development of complex coordination clusters.

If an acidified aqueous molybdate(VI)-containing solution is reacted with an appropriate reducing agent or directly with Mo^V centres (MoCl_5), the deep-blue ring-shaped clusters $\{\text{Mo}_{154}\}$ or $\{\text{Mo}_{176}\}$ are immediately formed in solution, in proportions that depend mainly on the pH. They can be precipitated under certain conditions. But if the larger cluster is kept in solution, we find that a molecular growth process occurs near its central cavity in the presence of an excess of molybdate (see Methods) and reducing agent such as L-ascorbic acid causing consecutive electron transfer⁹. Dark-blue crystals of the mixed-crystal compound $0.5\{\text{Mo}_{176}\}/0.5\{\text{Mo}_{248}\}/16\text{Na}^+/\sim 350\text{H}_2\text{O}$ (**1a**), containing the initially formed and known cluster $\{\text{Mo}_{176}\}$ ($\equiv \{\{\text{Mo}_{216}\}\{\text{Mo}_8\}_{16}\{\text{Mo}_{16}\}_{16}\}$) (**2**; refs 7, 8) and the final cluster species $\{\text{Mo}_{248}\} \equiv [\{\{\text{Mo}_{216}\}\{\text{Mo}_8\}_{16}\{\text{Mo}_{16}\}_{16}\}\{\text{Mo}_{36}\}_2]$ (**3**) in the ratio 1:1, precipitate in high yield. **2** is also abundant in crystals of the non-mixed-crystal

compound⁷ $\{\text{Mo}_{176}\}/16\text{Na}/\sim 400\text{H}_2\text{O}$ (**2a**) (the corresponding Li compound is isostructural and the $\{\text{Mo}_{176}\}$ cluster containing CH_3OH ligands has also the same structure⁸) and **3** is abundant in $\{\text{Mo}_{248}\}/16\text{Na}/\sim 250\text{H}_2\text{O}$ (**3a**)—for which only a few crystals in the presence of several mainly amorphous by-products could be obtained using a complicated time-consuming heterogeneous reaction with SnCl_2 as reducing agent (A.M., E. Krickemeyer, H.B. and M.S., unpublished results). In this context, note that **3** is less soluble in water than **2** which has a relatively higher density of H_2O ligands at the cluster surface. This leads, in the presence of both anions **2** and **3**, to an 'early' co-precipitation before all the cluster anions in solution show the growth process $2 \rightarrow 3$.

The formula of the mixed-valence compound **1a** (type III according to the well-known classification of M. B. Robin and P. Day¹²) was determined by elemental analyses, single-crystal X-ray structure analysis (including the calculation of bond valence sums in order to distinguish between (terminal) O, O(H) and OH_2 centres), redox titrations (for the determination of the formal number of Mo^{V} centres) and thermogravimetry (for the determination of the amount of crystal water).

The crystal structure analysis shows the abundance of the above-mentioned discrete ring-shaped cluster $\{\text{Mo}_{176}\}$ (**2**) and the 'extended' cluster $\{\text{Mo}_{248}\}$ (**3**) statistically distributed over the same positions in the crystal lattice in a ratio of $\sim 1:1$ (Fig. 1). This follows from the facts that: first, compounds **2a** and **3a** have practically the same lattice constants and show the same type of packing of the two kinds of clusters having the same diameter (the $\{\text{Mo}_{36}\}$ hub-caps are incorporated in such a way that the

packing of the wheels is not affected), and second, the electron density on the cluster positions of **1** corresponds to the weighted average of that of **2** and **3**. The formula obtained for **3** is $[\{\text{Mo}_2\}_{16}\{\text{Mo}_8\}_{16}\{\text{Mo}_1\}_{16}\{\text{Mo}_{36}\}_2]^{16-} = [\{\text{Mo}_2^{\text{VI}}\text{O}_5(\text{H}_2\text{O})_2\}_{16}\{\text{Mo}_8^{\text{VI/V}}\text{O}_{26}(\mu_3\text{-O})_2\text{H}(\text{H}_2\text{O})_3\text{Mo}^{\text{VI/V}}\}_{16}\{\text{Mo}_{36}^{\text{VI/V}}\text{O}_9(\text{H}_2\text{O})_{24}\}_2]^{16-}$.

The discrete extended cluster **3** shows, besides an identical arrangement of atoms in the peripheral ring region as in **2**, two additional polyoxomolybdate fragments which can be imagined as hub-caps of a wheel having the composition $\{\text{Mo}_{36}\text{O}_9(\text{H}_2\text{O})_{24}\}$ and which form also a new type of cavity (Fig. 2). The peripheral ring region is built up by the above-mentioned $\{\text{Mo}_2\}$ ($\equiv \{(\text{H}_2\text{O})(\text{O}_{\text{term}})_2\text{MoOMo}(\text{O}_{\text{term}})_2(\text{H}_2\text{O})\}$), $\{\text{Mo}_8\}$ and $\{\text{Mo}_1\}$ units as in **2**, each occurring 16 times. The terminal oxygen atoms (O_{term}) of the corresponding $\text{Mo}(\text{O}_{\text{term}})_2$ groups of $\{\text{Mo}_2\}$ fragments have a high electron density and therefore a high affinity for protonation. Those terminal oxygen atoms which are directed approximately towards the ring plane are the starting points for the observed subsequent growth and condensation process $2 \rightarrow 3$, thus leading to a 'closing' of the upper and lower aperture of the initially formed $\{\text{Mo}_{176}\}$ -type ring by (formally) uncharged polyoxomolybdate fragments $\{\text{Mo}_{12}\text{Mo}_{24}\text{O}_9(\text{H}_2\text{O})_{24}\}$ which do not exist in separate form (see below). As these hub-caps are (exclusively) connected to the above-mentioned 16 terminal oxygen atoms of the eight $\{\text{Mo}_2\}$ groups of **2** (refs 7, 8) on both the upper and the lower parts of the ring, they are accordingly turned at an angle of $360/16$ degrees relative to one another, in agreement with the corresponding relative positions of the $\{\text{Mo}_2\}$ groups.

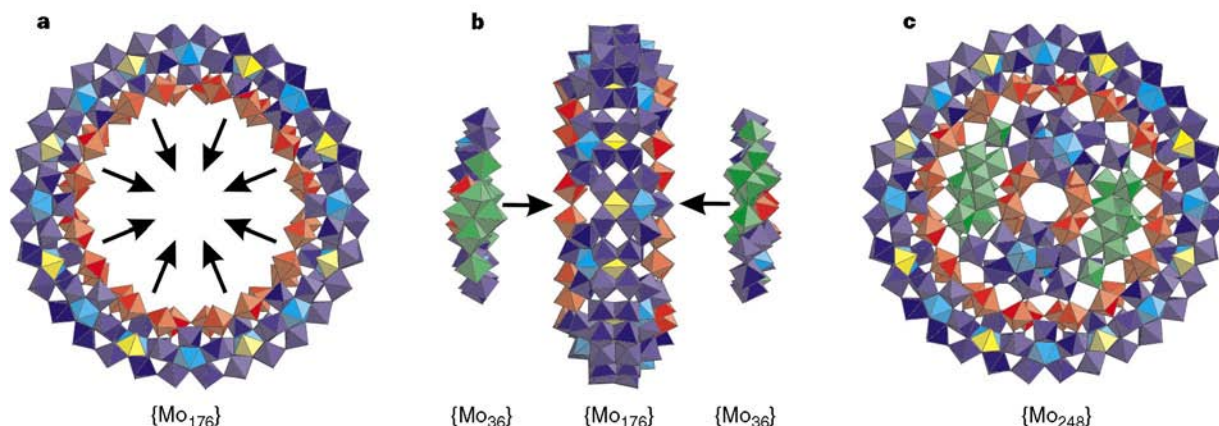


Figure 1 Schematic view of the relation between the $\{\text{Mo}_{176}\}$ and the $\{\text{Mo}_{248}\}$ clusters in polyhedral representation. Shown are the $\{\text{Mo}_{176}\}$ cluster (**a**, with arrows indicating the pathway of the growth process) and its side view (**b**), perpendicular to **a**, demonstrating the 'addition' of the two $\{\text{Mo}_{36}\}$ -type hub-caps

which finally results in the $\{\text{Mo}_{248}\}$ cluster **c**. Colour code for the building blocks: $\{\text{Mo}_8\}$ blue (central pentagonal bipyramids: turquoise), $\{\text{Mo}_8\}$ ' green, $\{\text{Mo}_2\}$ red, $\{\text{Mo}_1\}$ yellow.

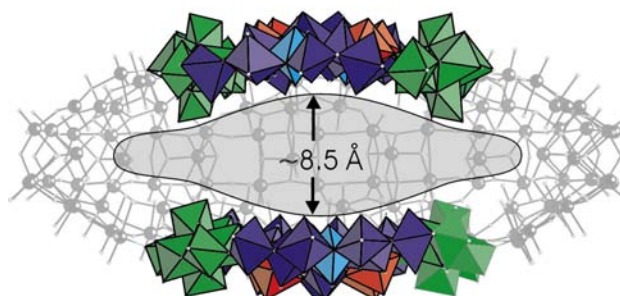


Figure 2 View into the cavity (shown grey) of the $\{\text{Mo}_{248}\}$ cluster. The hub-caps are in polyhedral representation, and the peripheral ring region is shown as a ball-and-stick model omitting the front part of the peripheral ring region.

The present reaction is of general importance for conservative self-organization processes with formation of complex structures from more simple ones. It can be concluded that in polyoxometalate systems a stepwise growth process can take place, whereby an initial cluster acts as a compartment for the aggregation of additional building units to form clusters that are not stable on their own (see also ref. 10). The aggregation of units within the cavity of the $\{\text{Mo}_{176}\}$ cluster can be regarded as a model both for crystal growth (especially for the initial nucleation process, which is still not understood) and also for metal-centre assembly processes in biological systems. In particular such growth processes might be allied to biomineralization in compartments or to metal-centre uptake in metal-storage proteins, such as the formation of the polyoxometalate nucleus of the molybdenum storage protein of the N_2 -fixing microorganisms *Azotobacter vinelandii*¹¹. □

Methods

Synthesis of 1a. An aqueous solution of 3 g of Na_2MoO_4 (12.5 mmol) in 45 ml H_2O is acidified with 2.5 ml of 32% HCl under continuous stirring. After adding 1.2 ml of ascorbic acid solution (0.5 M, 0.6 mmol), within a few minutes the reaction mixture turns yellow, then green, and finally dark blue. The reaction flask was covered with a stopper and allowed to stand for 4–5 weeks for crystallization (yield 0.38 g; correct analysis).

Selected physical data. Infrared (KBr pellet): ν (cm^{-1}) = 1,610 ($\text{m}, \delta(\text{H}_2\text{O})$), 976(w), 914(m) [ν (MoO_4)], 745(s), 629(m), 557(s); resonance-Raman (KBr pellet): ν (cm^{-1}) = 946 (m), 791(s), 540(m), 464(s), 327(s), 215(s); near-infrared/visible (in 40% MeOH/ H_2O for the estimation of the (formal) number of Mo^{V} centres): λ (nm) (ϵ_{M}) = 753 ($3.5 \times 10^5 \text{ l mol}^{-1} \text{ cm}^{-1}$), 1,016 ($2.5 \times 10^5 \text{ l mol}^{-1} \text{ cm}^{-1}$).

X-ray structure analysis. Space group $C222_1$, $a = 53.835(4)$, $b = 31.505(2)$, $c = 66.818(5)$ Å, $V = 113329(15)$ Å³, $Z = 4$, $\rho = 2.279 \text{ g cm}^{-3}$, $\mu = 2.38 \text{ mm}^{-1}$, $F(000) = 74,512$, crystal size = $0.28 \times 0.28 \times 0.10 \text{ mm}$. Crystals were removed from the mother liquor and immediately cooled to 183(2) K on a Bruker axis SMART diffractometer (Mo-K α , graphite monochromator). A total of 185,373 reflections ($1.83^\circ < \theta < 22.49^\circ$) were collected, of which 71,853 unique reflections ($R_{\text{int}} = 0.085$) were used. The structure was solved using the program SHELXS-97 and refined (3,333 parameters) using the program SHELXL-97 to $R = 0.0751$ for 35,731 reflections with $I > 2\sigma(I)$ (both programs from G. M. Sheldrick, Univ. Göttingen, 1997). The Na^+ ions, as well as several crystal water molecules, could not be localized due to their disorder.

Received 31 July; accepted 15 September 1998.

- Müller, A. & Pope, M. T. in *From Simplicity to Complexity Part II: Information—Interaction—Emergence* (eds Mainzer, K., Müller, A. & Saltzer, W. G.) 57–68 (Vieweg, Braunschweig, 1998).
- Pope, M. T. & Müller, A. Polyoxometalate chemistry: an old field with new dimensions in several disciplines. *Angew. Chem. Int. Edn Engl.* **30**, 34–48 (1991).
- Rhule, J. T., Hill, C. L., Judd, D. A. & Schinazi, R. F. Polyoxometalates in medicine. *Chem. Rev.* **98**, 327–357 (1998).
- Katsoulis, D. E. A survey of applications of polyoxometalates. *Chem. Rev.* **98**, 359–387 (1998).
- Müller, A. et al. $[\text{Mo}_{154}(\text{NO})_{10}(\text{OH})_{28}(\text{H}_2\text{O})_7]^{(25-)-}$: a water-soluble big wheel with more than 700 atoms and a relative molecular mass of about 24000. *Angew. Chem. Int. Edn Engl.* **34**, 2122–2124 (1995).
- Müller, A., Peters, F., Pope, M. T. & Gatteschi, D. Polyoxometalates: very large clusters—nanoscale magnets. *Chem. Rev.* **98**, 239–271 (1998).
- Müller, A. et al. Formation of a ring-shaped reduced ‘metal oxide’ with the simple composition $[(\text{MoO}_3)_{176}(\text{H}_2\text{O})_{80}\text{H}_{11}]$. *Angew. Chem. Int. Edn Engl.* **37**, 1220–1223 (1998).
- Müller, A., Koop, M., Böge, H., Schmidtman, M. & Beugholt, C. Exchanged ligands on the surface of a giant cluster: $[(\text{MoO}_3)_{176}(\text{H}_2\text{O})_{80}(\text{CH}_3\text{OH})_{17}\text{H}_{11}]^{(92-)-}$. *J. Chem. Soc., Chem. Commun.* 1501–1502 (1998).
- Williams, N. H. & Yandell, J. K. Outer-sphere electron-transfer reactions of ascorbate anions. *Aust. J. Chem.* **35**, 1133–1144 (1982).
- Müller, A. et al. Unusual stepwise assembly and molecular growth: $[\text{H}_{14}\text{Mo}_{37}\text{O}_{112}]^{14-}$ and $[\text{H}_3\text{Mo}_{57}\text{V}_6(\text{NO})_8(\text{H}_2\text{O})_{12}(\text{MoO}_4)_6]^{21-}$. *Chem. Eur. J.* **4**, 1000–1006 (1998).
- Müller, A. et al. Comparative *in-vivo* and *in-vitro* ^{99}Mo -time-differential-perturbed-angular-correlation studies on the nitrogenase MoFe protein and on other Mo species of different N_2 -fixing bacteria. *Eur. J. Biochem.* **246**, 311–319 (1997).
- Robin, M. B. & Day, P. Mixed valence chemistry - a survey and classification. *Inorg. Chem. Radiochem.* **10**, 247–422 (1967).

Acknowledgements. We thank E. Krickemeyer, B. Hauptfleisch, C. Beugholt, E. Diemann and F. Peters for their collaboration. This work was supported by the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie.

Correspondence and requests for materials should be addressed to A.M. (e-mail: a.mueller@uni-bielefeld.de). Further details of crystal structure determination may be obtained from the Fachinformationszentrum Karlsruhe, D-76344 Eggenstein-Leopoldshafen, Germany, on quoting the depository no. CSD 410267, names of authors and journal citation.

Energy landscapes of receptor–ligand bonds explored with dynamic force spectroscopy

R. Merkel*†, P. Nassoy*‡, A. Leung*, K. Ritchie* & E. Evans*§

* Departments of Physics and Pathology, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z1

† Physikdepartment der Technischen Universität München, 85748 Garching, Germany

‡ Physico-Chimie de l'Institut Curie, 75231 Paris, France

§ Biomedical Engineering, Boston University, Boston, Massachusetts 02215, USA

Atomic force microscopy (AFM)^{1,2} has been used to measure the strength of bonds between biological receptor molecules and their ligands^{3–6}. But for weak noncovalent bonds, a dynamic spectrum of bond strengths is predicted as the loading rate is altered, with the measured strength being governed by the prominent barriers traversed in the energy landscape along the force-driven bond-dissociation pathway⁷. In other words, the pioneering early AFM measurements represent only a single point in a continuous spectrum of bond strengths, because theory predicts that these will depend on the rate at which the load is applied. Here we report the strength spectra for the bonds between streptavidin (or avidin) and biotin⁸—the prototype of receptor–ligand interactions used in earlier AFM studies^{3–5}, and which have been modelled by molecular dynamics^{9,10}. We have probed bond formation over six orders of magnitude in loading rate, and find that the bond survival time diminished from about 1 min to 0.001 s with increasing loading rate over this range. The bond strength, meanwhile, increased from about 5 pN to 170 pN. Thus, although they are among the strongest noncovalent linkages in biology (affinity of 10^{13} to 10^{15} M^{-1})^{8,11}, these bonds in fact appear strong or weak depending on how fast they are loaded. We are also able to relate the activation barriers derived from our strength spectra to the shape of the energy landscape derived from simulations of the biotin–avidin complex.

To measure strengths, we used two modes of a biomembrane force probe (BFP)¹² as described in Fig. 1 legend: a vertical mode with high resolution in position and force (2–5 nm and 0.2–0.5 pN) for testing weak bonds under slow loading (Fig. 2a); and a horizontal mode with modest resolution (8–10 nm and 1–10 pN) for testing strong bonds under fast loading (Fig. 2b). The BFP tip and test surface were prepared with a paucity of reactive sites to limit the frequency of bond formation to 1 per 7–10 touches in the tests (see Methods). As such, the likelihood of forming and breaking single bonds was expected to be greater than 0.9. To obtain sufficient numbers (50–100) of rupture forces at each loading rate, several hundred cycles of approach–touch–separation were performed by computer-controlled piezo displacement of either the transducer or the test surface. The nominal loading rate k_{FV} was preselected by setting the BFP force constant k_{F} in the range 0.1–3 pN nm^{−1} and piezo retraction speed v_{r} in the range 1–20,000 nm s^{−1}. Force histograms were compiled at 11 and 8 loading rates between 0.05 to 60,000 pN s^{−1} for biotin–avidin and biotin–streptavidin, respectively. A montage of histograms is shown in Fig. 3a for biotin–streptavidin. Complete spectra of the most frequent rupture forces f^* versus $\log r_{\text{f}}$ (where r_{f} = loading rate) are plotted in Fig. 3b.

To understand why strength depends on loading rate, it is important to recognize that the lifetime of a bond sustained by weak noncovalent interactions diminishes rapidly when subjected to force because of thermal activation. Conceptually, the energy landscape is tilted by force (Fig. 4a), which lowers energy barriers,

decreases the likelihood of bond survival, and speeds up dissociation. So, we might expect the form of a strength spectrum obtained under rising force in probe tests to be complicated and difficult to interpret. However, when linear regimes appear over many orders of magnitude of loading rate as in Fig. 3b, interpretation of the spectrum is simple: each regime produces an image of a sharp energy barrier at a fixed location along the unbinding pathway⁷. The energy contour in configuration space local to a sharp barrier (called the transition state) is highly curved and therefore does not change shape as the barrier height falls under rising force f (Fig. 4a). The thermally averaged displacement in configuration space needed to reach the top of the barrier does not shift with force and the displacement maps to a constant position x_β along the direction of force. First postulated intuitively by Bell¹³, lowering the barrier by the mechanical potential fx_β leads to exponential amplification of the dissociation kinetics, that is, off rate $\nu \approx \nu_0 \exp(fx_\beta/k_B T)$. Hence, the relevant force scale $f_\beta = k_B T/x_\beta$ for thermally activated rupture is thermal energy ($k_B T \approx 4.1 \times 10^{-21}$ J or ≈ 4.1 pN nm at room temperature) divided by the projected bond displacement, not the maximum gradient in an energy landscape. The force statistics in probe tests are predicted by a first-order kinetic process where dissociation rate increases rapidly with the rising force⁷. For a single barrier, the peak f^* in the force distribution shifts to higher force in proportion to $\log_e$ (loading rate) with a slope f_β . Much less trivial,

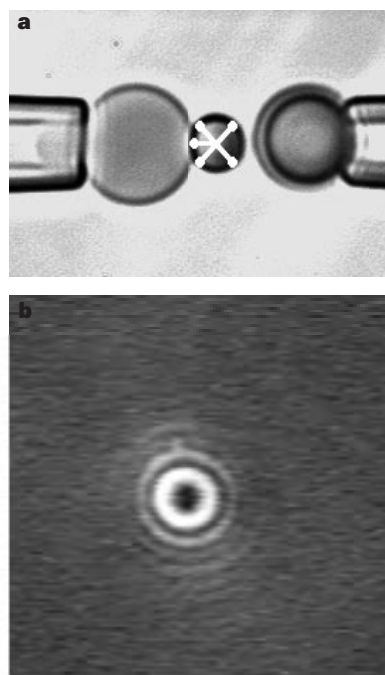


Figure 1 The spring in the biomembrane force probe BFP is a pressurized membrane capsule¹². Membrane tension sets the force constant k_t (force/capsule extension) and is controlled by micropipette suction P and radius R_p , $k_t \sim PR_p$. Using a red blood cell as the transducer, the BFP stiffness was tuned between 0.1 and 3 pN nm⁻¹ to measure forces from 0.5 to 1,000 pN. As the BFP tip, a glass microbead of 1–2 μ m diameter was chemically glued to the membrane (see Methods). **a**, Operated on the stage of an inverted microscope, the BFP (on the left) in the horizontal mode was kept stationary and the microbead test surface (on the right) was translated to/from contact with the BFP tip by precision piezo control. With fast video ($\sim 1,000$ frames per s) processing, a simulated cursor was required to track the image of the bead as shown, which yielded a resolution of 8–10 nm for transducer deflection. **b**, Reflection interference contrast image of the BFP tip translated along the optical axis by piezo control to/from a coverglass test surface in the vertical mode. Standard video (30 frames per s) processing of the circular interference pattern was used to track elevation of the tip at a resolution of 2–5 nm. Transducer deflection was obtained from the difference between piezo translation and bead displacement.

complex macromolecular bonds involve many interactions that create a mountainous terrain of barriers in the energy landscape. Assuming a cascade of sharp barriers, the strength spectrum is predicted to follow a piece-wise continuous sequence of linear regimes with ascending slopes⁷. The abrupt increase in slope from one regime to the next signifies that an outer barrier has been suppressed by force and that an inner barrier becomes the dominant kinetic impedance, as shown in Fig. 4a. The regime governed by a particular barrier spans a range of $\log_e$ (loading rate) determined by its height relative to adjacent barriers (see Methods). The off rate rises as a staircase of exponentials in force that amplify off rate less and less from one to the next.

Applying these concepts to the spectra plotted in Fig. 3b, we derive locations of prominent energy barriers that govern strength of biotin–streptavidin and biotin–avidin bonds. How does this one-dimensional map along the direction of probe force compare with detailed molecular dynamics (MD) simulations of biotin–(strept)avidin interactions? In separate simulations, biotin was extracted from a binding pocket of streptavidin⁹ and avidin¹⁰ by pulling on the outer end with a pseudo-mechanical spring. Chemically and structurally, the binding pockets in avidin and streptavidin are very similar except biotin forms an additional nonpolar interaction and three additional hydrogen bonds in avidin^{14,15}. Also, the longer ‘3–4’ loop in avidin seems to close more tightly behind biotin in the bound state^{14–16}. Revealing the inherent molecular complexity, the simulations yield a dynamic superposition of many polar (hydrogen bonds and water bridges) and nonpolar (to aromatic residues) interactions along the unbinding trajectories, which are emphasized quite differently in each report^{9,10}. Even so, common qualitative features are described that provide important clues to the thermally averaged free energy landscape relevant on laboratory timescales. First, within an initial displacement of less than 0.2 nm, unbinding began with detachment of the ‘head’ (ureido ring) of

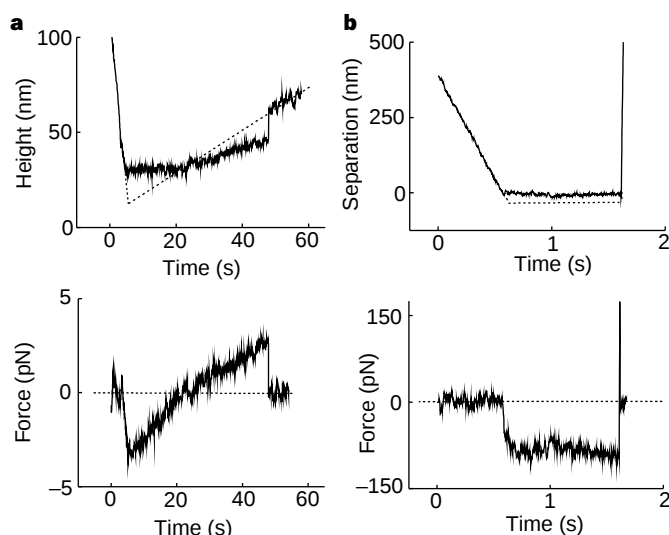


Figure 2 BFP tip–substrate distance and force versus time for cycles of approach–touch–separation with formation and rupture of a bond. **a**, Loaded at extremely slow rate, a bond held the tip to the surface for ~ 24 s and broke at ~ 3 pN as the piezo retracted the transducer (dashed trajectory). The fluctuations in tip position were due to thermal excitations of the BFP (mean square displacement $\sim k_B T/k_t$). Stretch of the PEG polymers that linked the bond to the glass surfaces is shown by the slight upward movement (~ 15 nm) under force before detachment. Because of polymer compliance, the true loading rate felt by a bond at nominal rates ($k_t v_t$) below 10 pN s⁻¹ had to be obtained from the force versus time. **b**, Loaded at extremely fast rate, a bond held the tip to the surface for ~ 0.003 s (spike in force) and broke at ~ 170 pN as the piezo retracted the test surface (dashed trajectory). The force fluctuations were due to position uncertainties $\times$ BFP stiffness.

biotin from a nest of hydrogen bonds, water bridges and nonpolar interactions deep in the binding pocket. Next, forces reached maximal values followed by sudden displacements of biotin at a distance of ~ 0.5 nm in the biotin-streptavidin simulation (attributed to rupture in a transient network of water bridges and hydrogen bonds) and at ~ 0.4 nm in the biotin-avidin simulation (attributed both to polar and to nonpolar interactions). Finally, as biotin left the pocket, a prominent jump occurred with lower forces at ~ 1 nm in both simulations (attributed to hydrogen bonds) and biotin was observed still to cling to peripheral polar groups at ~ 1.4 nm in avidin simulations. To show this behaviour clearly, we have plotted a record of the instantaneous interaction energies between biotin and avidin calculated over a half-nanosecond time course of extraction in the simulations of Israilev *et al.*¹⁰ (Fig. 4b).

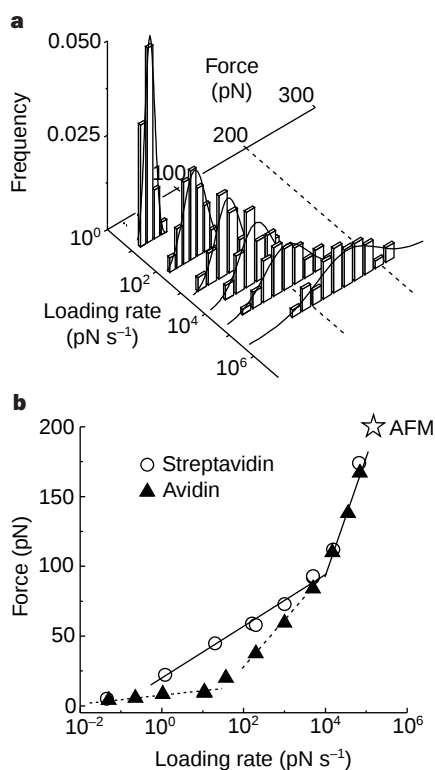


Figure 3 Biotin-streptavidin bond strengths. **a**, Force histograms from tests of single biotin-streptavidin bonds demonstrate shift in peak location and increase in width with increase in loading rate. Gaussian fits used to determine the most frequent rupture force or bond strength are shown. Governed ideally by the thermal force f_b , standard deviations σ_f of the distributions also reflected uncertainties in position Δx and video sampling time Δt_v , that is, $\sigma_f \sim [f_b^2 + (k_f \Delta x)^2 + (r_f \Delta t_v)^2]^{1/2}$. As σ_f increased from ± 1 pN at the slowest rate to ± 60 pN at the fastest rate, the standard error in mean force (the statistical measure for error in strength) ranged from ± 0.3 pN to ± 5 pN. **b**, Dynamic strength spectra for biotin-streptavidin (circles) and biotin-avidin (triangles) bonds. Defined as thermal energy $k_B T$ ÷ distance x_b , the slopes (f_b) of the solid lines in the biotin-streptavidin spectrum map activation barriers at $x_b \approx 0.5$ nm and 0.12 nm along the direction of force based on values of 8 pN and 34 pN. Merging with biotin-streptavidin above 85 pN, the high-strength regime for biotin-avidin also maps an inner barrier at $x_b \approx 0.12$ nm but the slope $f_b \approx 13$ – 14 pN of the intermediate strength regime (dashed line) between 38 pN and 85 pN indicates that the next barrier maps to $x_b \approx 0.3$ nm. Slight curvature and reduction in slope between 38 pN and 11 pN suggests that the barrier extends to ~ 0.5 nm. Below 11 pN, the biotin-avidin spectrum exhibits a low-strength regime (dashed line) with a slope of $f_b \approx 1.4$ pN that maps to $x_b \approx 3$ nm. Consistent with the high-strength regime is the biotin-streptavidin strength ($\star_{AFM}$) measured recently by atomic-force microscopy (AFM) using a carbon nanotube as the tip²⁴ and the biotin-avidin strength (not shown) measured previously⁴ by AFM.

Transition states are readily identified by regions with a paucity of states where biotin passes quickly. Marked in Fig. 4b, the activation barriers derived from the high and intermediate strength regimes in Fig. 3b correlate with regions of rarified statistics and the qualitative appearance of the energy landscape. The interpretation is that the transition states implied by these features persist on long timescales and that the molecular reaction coordinate perhaps deviates by ~ 40 – 45° from the direction of force local to the second transition state. But surprisingly, the outer barrier indicated by the low-strength regime in Fig. 3b is 2–3-fold more distant than the last transition state seen in the MD simulation.

Guided by the MD simulation, we expect the outer barrier to emanate from molecular interactions with the flexible ‘3–4’ loop, which closes behind biotin in crystallographic images^{14–16} of the

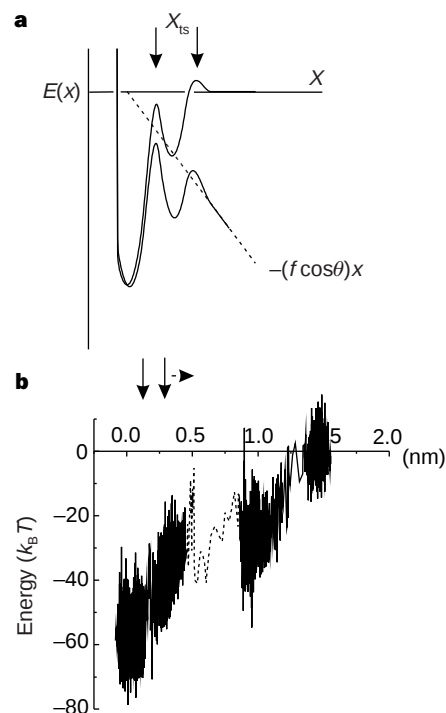


Figure 4 Conceptual and real (MD) energy landscapes traversed along a molecular reaction coordinate under force. **a**, Oriented at an angle θ to the molecular coordinate x , external force f adds a mechanical potential, $-(f \cos \theta)x$, that tilts the landscape and lowers barriers. The inner barrier emerges to dominate kinetics when the outer barrier falls to a level $\geq k_B T$ below it under force. For sharp barriers, the local energy contours, called transition states, are highly curved and change little in shape or location under force. Even though the unbinding pathway may be tortuous and the orientation fluctuates wildly, the energy weighted locations of sharp barriers can project as constant distances $x_b = \langle x_{ts} \cos \theta \rangle$ along the direction of force. **b**, Instantaneous interaction energy between biotin and avidin computed along a half-nanosecond extraction from the binding pocket in the simulations of Israilev *et al.*¹⁰ (kindly provided to us by K. Schulten and co-workers, University of Illinois). Bordered by regions of rapid intense fluctuations, locations of rarified statistics reveal transition states expected in a thermally averaged free energy landscape⁷. Arrows mark barrier locations derived from the strength spectrum for biotin-avidin in Fig. 3b.

bound state and was set in an open conformation in MD simulations. Consistent with this expectation, mutations that delete the '3-4' loop in streptavidin result in major reductions in magnitude of binding enthalpy^{16,17}. Interestingly, in the absence of biotin, the '3-4' loop disappears in crystallographic images indicating that the loop becomes disordered and flexible. Moreover, although not currently implicated in biotin-avidin binding, other longer loops also border the channel that leads to the binding pocket. Thus, our speculation is that the outer barrier at ~3 nm represents interactions on laboratory timescales of the spacer-linked biotin with soft, flexible elements well beyond the binding pocket. Even though biotin-(strept)avidin bonds can break under very small forces, the location of the outer barrier at ~3 nm leads to a significant difference in energy between the outer and nearby inner barriers as indicated by the difference in log/loading rate) intercepts of the low and intermediate strength regimes.

The profound effect of thermal activation on strength of non-covalent linkages in biology has been recognized by researchers studying cell adhesion dynamics in shear flow^{18,19} and recently also in the unfolding of tandem immunoglobulin-like domains in long proteins^{20,21}. What is not well known, however, is that thermal activation in a complex biomolecular assembly is likely to be governed by a rugged energy landscape with more than one kinetic barrier. We have shown here that to explore such a landscape with force probes, experiments have to be performed over an enormous range of loading rates. □

Methods

Chemical preparation of BFP tips and test surfaces. First, amino silane (AEAPTMS, United Chemical Technologies, PA) groups were covalently bound to glass microbeads and coverslips. Next, amine-reactive polyethylene oxide polyethylene glycol (PEG) polymers with and without biotin end groups (mixture of NHS-PEG3400-VS and NHS-PEG3400-biotin, Shearwater Polymers, AL) were covalently linked to the silanized surfaces. Finally, the biotinylated beads and coverglasses were exposed to excess (strept)avidin and then washed. Even though almost completely saturated with (strept)avidin, the surfaces still had a number of free biotin groups. Thus, bonds formed very infrequently when identically prepared tip and test surfaces were touched together. (No bonding was detected when the PEG polymers on the test surface were terminated with methyl groups or when free biotin was blown at the tip and test surface by an auxiliary micropipette before touch.) A red cell covalently linked with PEG-biotin polymers was pushed together with an avidinated microbead to construct the probe as shown in Fig. 1a.

Analysis of strength spectra. Slopes of linear regimes in strength versus log_e(loading rate) map energy barriers to fixed distances x_β along the direction of force⁷. Each slope is the force scale $f_\beta = k_B T/x_\beta$ for e-fold amplification of dissociation rate impeded by a particular barrier. Barriers emerge in succession from outer to inner positions to dominate kinetics. Reflecting the Arrhenius dependence in the off rate, $v_0 = (1/t_D) \exp(-E_b/k_B T)$, differences in logarithmic intercept $\log_e(r_f)_{f=0}$ and slope f_β of regimes in Fig. 3b expose differences in barrier heights, $\Delta E_b \approx k_B T \{\Delta \log_e[f_\beta/t_D] - \Delta \log_e(r_f)_{f=0}\}$, within an unknown variation in a diffusive relaxation time t_D . Worked out by Kramers^{22,23}, the frequency $1/t_D$ in liquids is governed by viscous friction γ_f and a product of length scales $l_{a,ts}$ ($l_a \equiv$ confinement length in the bound state and $l_{ts} \equiv$ impedance width of the transition state), that is, $1/t_D \approx k_B T/(\gamma_f l_{a,ts})$ and $k_B T/\gamma_f$ defines a diffusivity. From MD simulations⁹, values of $\gamma_f \sim 2 \times 10^{-8}$ pN s nm⁻¹ and $x_\beta x_{ts} \sim 0.01$ – 0.1 nm² imply that $1/t_D$ is $\sim 10^9$ – 10^{10} s⁻¹ and that log_e(f_β/t_D) is ~ 21 – 25 for a rate scale in pN s⁻¹.

Received 5 May; accepted 5 October 1998.

1. Binnig, G., Quate, C. F. & Gerber, C. H. Atomic force microscopy. *Phys. Rev. Lett.* **56**, 930–933 (1986).
2. Drake, B. et al. Imaging crystals, polymers, and processes in water with the atomic force microscope. *Science* **243**, 1586–1589 (1989).
3. Lee, G. U., Kidwell, D. A. & Colton, R. J. Sensing discrete streptavidin-biotin interactions with atomic force microscopy. *Langmuir* **10**, 354–357 (1994).
4. Florin, E.-L., Moy, V. T. & Gaub, H. E. Adhesive forces between individual ligand receptor pairs. *Science* **264**, 415–417 (1994).
5. Moy, V. T., Florin, E.-L. & Gaub, H. E. Intermolecular forces and energies between ligands and receptors. *Science* **264**, 257–259 (1994).

6. Hinterdorfer, P., Baumgartner, W., Gruber, H. J., Schilcher, K. & Schindler, H. Detection and localization of individual antibody-antigen recognition events by atomic force microscopy. *Proc. Natl Acad. Sci. USA* **93**, 3477–3481 (1996).
7. Evans, E. & Ritchie, K. Dynamic strength of molecular adhesion bonds. *Biophys. J.* **72**, 1541–1555 (1997).
8. Green, N. M. Avidin. *Adv. Protein Chem.* **29**, 85–133 (1975).
9. Grubmüller, H., Heymann, B. & Tavan, P. Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science* **271**, 997–999 (1996).
10. Izrailev, S., Stepanians, S., Balsera, M., Oono, Y. & Schulten, K. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys. J.* **72**, 1568–1581 (1997).
11. Chilkoti, A. & Stayton, P. S. Molecular origins of the slow streptavidin-biotin dissociation kinetics. *J. Am. Chem. Soc.* **117**, 10622–10628 (1995).
12. Evans, E., Ritchie, K. & Merkel, R. Sensitive force technique to probe molecular adhesion and structural linkages at biological interfaces. *Biophys. J.* **68**, 2580–2587 (1995).
13. Bell, G. I. Models for the specific adhesion of cells to cells. *Science* **200**, 618–627 (1978).
14. Weber, P. C., Ohlendorf, D. H., Wendoloski, J. J. & Salemme, F. R. Structural origins of high-affinity biotin binding to streptavidin. *Science* **243**, 85–88 (1989).
15. Livnah, O., Bayer, E. A., Wilchek, M. & Sussman, J. L. Three-dimensional structures of avidin and the avidin-biotin complex. *Proc. Natl Acad. Sci. USA* **90**, 5076–5080 (1993).
16. Freitag, S., Le Trong, I., Klumb, L., Stayton, P. S. & Stenkamp, R. E. Structural studies of the streptavidin binding loop. *Protein Sci.* **6**, 1157–1166 (1997).
17. Chu, V., Freitag, S., Le Trong, I., Stenkamp, R. E. & Stayton, P. S. Thermodynamic and structural consequences of flexible loop deletion by circular permutation in the streptavidin-biotin system. *Protein Sci.* **7**, 848–859 (1998).
18. Alon, R., Hammer, D. A. & Springer, T. A. Lifetime of the P-selectin-carbohydrate bond and its response to tensile force in hydrodynamic flow. *Nature* **374**, 539–542 (1995).
19. Brunk, D. K., Goetz, D. J. & Hammer, D. A. Sialyl Lewis^x/E-selectin-mediated rolling in a cell-free system. *Biophys. J.* **71**, 2902–2907 (1996).
20. Rief, M., Gautel, M., Osterhelt, F., Fernandez, J. M. & Gaub, H. E. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* **276**, 1109–1112 (1997).
21. Oberhauser, A. F., Marszalek, P. E., Erickson, H. P. & Fernandez, J. M. The molecular elasticity of the extracellular matrix protein tenascin. *Nature* **393**, 181–185 (1998).
22. Kramers, H. A. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica (Utrecht)* **7**, 284–304 (1940).
23. Hanggi, P., Talkner, P. & Borkovec, M. Reaction-rate theory: fifty years after Kramers. *Rev. Mod. Phys.* **62**, 251–342 (1990).
24. Wong, S. S., Joselivich, E., Woolley, A. T., Cheung, C. L. & Lieber, C. M. Covalently functionalized nanotubes as nanometre-sized probes in chemistry and biology. *Nature* **394**, 52–55 (1998).

Acknowledgements. We thank A. Chilkoti, C. Cantor and the group of K. Schulten for helpful discussions. The work was supported by USPHS National Institutes of Health, Medical Research Council of Canada, and the Canadian Institute for Advanced Research Program in Science of Soft Surfaces and Interfaces.

Correspondence and request for materials should be addressed to E.E. (e-mail: evans@physics.ubc.ca).

The fate of subducted basaltic crust in the Earth's lower mantle

Kei Hirose*, Yingwei Fei, Yanzhang Ma & Ho-Kwang Mao

Geophysical Laboratory and Center for High Pressure Research, Carnegie Institution of Washington, 5251 Broad Branch Road NW, Washington DC 20015, USA

* Present address: Department of Earth and Planetary Sciences, Tokyo Institute of Technology, Ookayama, Tokyo 152-8551, Japan

The subduction of oceanic lithosphere into the Earth's deep interior is thought to drive convection and create chemical heterogeneity in the mantle. The oceanic lithosphere as a whole, however, might not subduct uniformly: the fate of basaltic crust may differ from that of the underlying peridotite layer because of differences in chemistry, density and melting temperature. It has been suggested that subducted basaltic crust may in fact become buoyant at the mantle's 660-km discontinuity, remaining buoyant to depths of at least 800 km, and therefore might be gravitationally trapped at this boundary to form a garnetite layer^{1,2}. Here we report the phase relations and melting temperatures of natural mid-ocean ridge basalt at pressures up to 64 GPa (corresponding to ~1,500 km depth). We find that the former basaltic crust is no longer buoyant when it transforms to a perovskite lithology at about 720 km depth, and that this transition boundary has a positive pressure-temperature slope, in contrast to the negative slope of the transition boundary in peridotite. We therefore predict that basaltic crust with perovskite lithology would gravitationally sink into the deep mantle. Our melting data suggest that, at the base of the lower mantle, the former basaltic

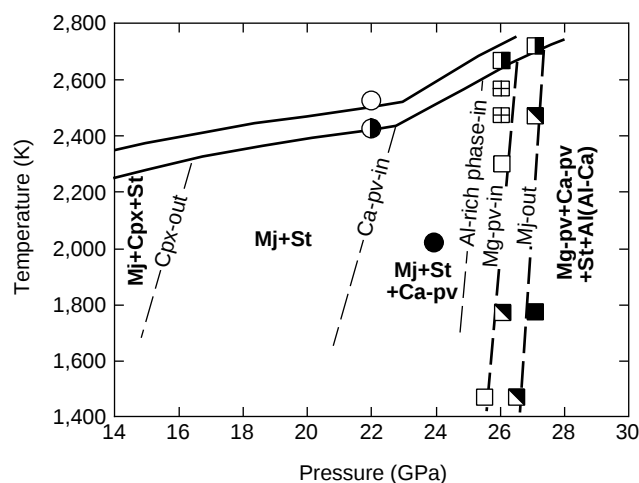


Figure 1 Phase relations in MORB composition up to 27 GPa. Solid lines represent solidus and liquidus temperatures. Lithology changes from garnetite to perovskitite between 25.5 GPa and 26.5 GPa (heavy dashed lines). Melting curve below 20 GPa is from Yasuda *et al.*¹², and sub-solidus phase relations below 25 GPa are from Irifune and Ringwood¹. Phases present in the experiments are magnesium-iron silicate perovskite (Mg-pv), majorite (Mj), Ca-perovskite (Ca-pv), aluminocalcic phase (Al-Ca), aluminous phase (Al), stishovite (St) and clinopyroxene (Cpx). Solid circle represents mineral assemblage of Mj+St+Ca-pv; open circle is above liquidus; open squares represent assemblage of Mj+St+Ca-pv+Al; squares with crosses represent assemblage of Mj+St+Ca-pv+Al+Ca; solid square represents assemblage of Mg-pv+St+Ca-pv+Al. The half-filled symbols represent mixed assemblages. At 26 GPa, Ca-pv is the liquidus phase, followed by Mj, St, and Al-Ca. At 27 GPa, Ca-pv is the liquidus phase, followed by St, Al-Ca, and Mg-pv.

crust would be partially molten if temperatures there were to exceed 4,000 K.

Sub-solidus phase relations in a mid-ocean-ridge basalt (MORB) composition were determined using a multi-anvil apparatus³ up to 27 GPa (Fig. 1). The compositions and proportions of constituent minerals at 27 GPa are presented in Table 1. A mineral assemblage of majorite + stishovite + CaSiO₃-perovskite was observed at 24 GPa and 2,023 K, consistent with the results of Irifune and Ringwood¹. We also confirmed the existence of an aluminous phase (Al-phase) with the calcium ferrite-type structure¹ at pressures higher than 24 GPa. In addition, a new aluminocalcic phase (Al-Ca-phase) was found in MORB composition at temperatures greater than 2,473 K and a pressure of 26 GPa, which may be related to the CAS phase⁴.

Although Irifune and Ringwood¹ only obtained a majorite-bearing assemblage and did not observe the Al-bearing Mg-perovskite formation in MORB composition up to 28 GPa, we found that the majorite-perovskite transformation starts at 26 GPa (equivalent to a depth of about 720 km) and 2,000 K with a small positive pressure-temperature slope (+0.0008 GPa K⁻¹) and completes within a 1 GPa interval. This discrepancy can be attributed to uncertainty in their pressure calibration because there is no reference point at pressures above the ilmenite-perovskite transition pressure (~23 GPa) at high temperatures. The fact that we obtained a perovskite-bearing assemblage in a MORB composition indicates that we have achieved a higher maximum pressure than did Irifune and Ringwood. The majorite-perovskite coexisting field is relatively narrow (less than 1 GPa) compared to that of the majorite-perovskite transformation in the system MgSiO₃-Mg₃Al₂Si₃O₁₂ (refs 5, 6). Because MORB has high Si, Al, Fe and Na contents, its mineralogy at mantle pressures is substantially more complicated than that of mantle peridotite, which consists solely of Mg-perovskite, CaSiO₃-perovskite and magnesiowüstite in the lower mantle⁷. The formation of the Al-phase and the Al-Ca-phase at the majorite-perovskite transformation pressure is directly related to the change in solubility of Al₂O₃ and Na₂O in Mg-perovskite and majorite as a function of pressure and temperature. The higher Al₂O₃ content of MORB also results in a higher majorite-perovskite transition pressure in MORB relative to that in peridotite.

Former basaltic crust has a higher density than the surrounding pyrolytic mantle above the 660-km discontinuity marked by perovskite formation in pyrolite (Fig. 2). It is less dense than the surrounding mantle just below the 660-km discontinuity. Once it transforms to perovskite lithology at about 720 km depth, its density significantly increases and it becomes denser than the surrounding pyrolytic mantle (Fig. 2). The zero-pressure density of basaltic crust with perovskite lithology is 4.23 g cm⁻³, based on X-ray diffraction data and the mineral proportions calculated by mass balance at 27 GPa and 2,473 K (Table 1), which is higher than that of pyrolytic mantle by +0.1 g cm⁻³. The calculated zero-pressure densities at 24 and 26 GPa are 3.87 and 3.92 g cm⁻³, respectively, consistent with previous results¹. The present results show that the former oceanic crust is buoyant only at depths between 660 km and 720 km. If the slabs accumulate to form a 'megalith'² of sufficient thickness (>60 km) underneath the 660-km discontinuity, then the density of basaltic crust will increase because of the transformation to perovskite lithology, and this transformation could drive slab penetration into the deep mantle.

Our present results indicate that the transformation to perovskite lithology in MORB occurs at a depth (720 km) that is much

Table 1 Compositions of starting material and solid phases at 27 GPa

	Mg-pv	Mj	St	Ca-pv	Al-phase	Al-Ca	SM
SiO ₂	37.38(53)	41.51(47)	96.82	43.20(47)	26.96(38)	40.25(34)	49.64
TiO ₂	3.58(14)	2.02(13)	0.11	1.56(03)	0.92(03)	0.18(06)	1.64
Al ₂ O ₃	16.14(64)	19.79(99)	1.26	4.49(62)	32.78(54)	40.95(25)	14.88
FeO*	23.01(85)	16.50(09)	0.46	3.45(43)	16.19(52)	1.00(18)	11.43
MnO	0.25(09)	0.08(01)	0.00	0.22(08)	0.05(01)	0.01(02)	0.18
MgO	17.85(35)	11.44(04)	0.38	0.89(19)	10.74(12)	3.46(15)	8.51
CaO	1.65(38)	3.75(07)	0.36	39.50(99)	1.87(04)	11.92(21)	10.55
Na ₂ O	0.87(17)	5.09(46)	0.16	0.68(15)	11.15(17)	1.30(05)	2.90
K ₂ O	0.02(03)	0.05(01)	0.02	0.12(02)	0.00(00)	0.29(02)	0.12
Cr ₂ O ₃	0.02(03)	0.15(02)	0.02	0.04(02)	0.10(01)	0.07(03)	-
Total	100.77	100.37	99.60	94.15	100.75	99.43	99.85
Proportion (wt %)	27	3	22	23	25		

Chemical composition analyses for Mg-perovskite (Mg-pv), majorite (Mj), stishovite (St), Ca-perovskite (Ca-pv) and Al-phase were obtained from a run at 27 GPa and 2,473 K, and the analysis for Al-Ca-phase (Al-Ca) was from a run at 27 GPa and 2,723 K. Starting material (SM) used in the experiments was a MORB glass¹². Phase identification was confirmed by micro-Raman and X-ray diffraction. Mineral proportions were calculated on the basis of mass balance using the measured compositions of minerals present and confirmed by electron microprobe composition maps. Zero-pressure densities were calculated using X-ray diffraction data obtained in this study, combining with the measured compositions and the calculated mineral proportions. The measured lattice parameters for minerals in the assemblage at 27 GPa are $a = 4.8067(4)$ Å, $b = 5.0089(4)$ Å, and $c = 7.0407(5)$ Å for Mg-perovskite; $a = 4.1783(2)$ Å and $c = 2.6716(2)$ Å for stishovite; and $a = 8.7356(7)$ Å, $b = 10.1283(3)$ Å, and $c = 2.8299(2)$ Å for Al-phase based on the calcium ferrite type-structure. The density of 4.25 g cm⁻³ for Ca-perovskite²⁵ is used in the calculations. The calculated bulk zero-pressure density at 27 GPa is 4.23 g cm⁻³. The measured lattice parameters at 26 GPa are $a = 11.5172(7)$ Å for majorite; $a = 4.1741(4)$ Å and $c = 2.6640(4)$ Å for stishovite; and $a = 8.7051(8)$ Å, $b = 10.1425(5)$ Å, and $c = 2.8213(3)$ Å for Al-phase. The calculated bulk zero-pressure density at 26 GPa is 3.92 g cm⁻³. The measured lattice parameters at 24 GPa are $a = 11.5758(8)$ Å for majorite; and $a = 4.1750(4)$ Å and $c = 2.6649(2)$ Å for stishovite. The calculated bulk zero-pressure density at 24 GPa is 3.87 g cm⁻³. Values in parentheses represent the standard deviation in the last digits.

* Total Fe as FeO.

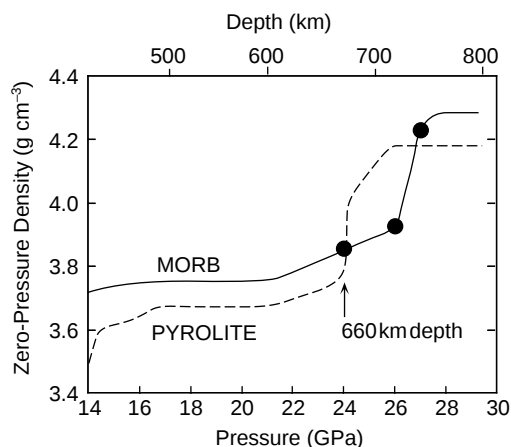


Figure 2 Comparison of zero-pressure density changes in MORB (solid line) and pyrolite (dashed line). Solid circles represent the calculated densities at 24, 26 and 27 GPa from X-ray diffraction and microprobe data. The density profile of pyrolite is from a previous study¹⁰. Pyrolite becomes denser than MORB at 660 km depth because of the transformation to perovskite lithology, but once MORB transforms to perovskite at 720 km depth, it is no longer buoyant in the deep mantle.

shallower than that expected^{1,2,8,9}. The transition boundary has a positive pressure–temperature slope, whereas the transition boundary in the underlying harzburgite has a negative slope¹⁰. Within a cool slab (for example 1,000 °C at the 660 km depth), the transformation to perovskite lithology in the basaltic crust and the underlying harzburgite layer would occur at a similar depth, implying that the delamination of the basaltic crust from the slab is unlikely at the 660 km discontinuity.

Melting experiments were carried out using a multi-anvil apparatus³ and laser-heated diamond cell¹¹ in a pressure range between 16 and 64 GPa (Fig. 3). The solidus temperatures are 2,400 K and 2,700 K at 22 GPa and 27 GPa, respectively, based on the multi-anvil experiments. The solidus phase assemblage at 27 GPa is Al-bearing Mg–perovskite and Ca–perovskite, stishovite, Al-phase and trace majorite (<3%). No important phase transformations have been reported at higher pressures up to 100 GPa (ref. 9), suggesting that this assemblage remains in the deep mantle, except that majorite would be completely transformed to perovskite at pressures slightly above 27 GPa. Analysis of run products above the solidus temperature at 27 GPa showed that Ca–perovskite is the liquidus phase, followed by stishovite, Al–Ca-phase and Al-bearing Mg–perovskite. The partial melt composition is enriched in MgO and FeO and depleted in SiO₂ and Al₂O₃, reflecting that Mg–perovskite is eliminated first near the solidus temperature. The partial melt is probably denser than the solid residue in MORB composition at lower mantle pressures because of its higher iron content. The melting curve of MORB determined in the laser-heated diamond cell is consistent with that determined by multi-anvil experiments for the same starting material between 22 and 27 GPa (Fig. 3), and also with previous results at low pressures¹². It is substantially lower than that of each constituent mineral: Mg–perovskite^{13,14}, Ca–perovskite^{14,15} and stishovite¹⁴. The present measurements of the melting temperature of MgSiO₃ are consistent with previous studies^{13,14,16} at 12–28 GPa.

The melting temperature of basalt is about 250 K lower than that of mantle peridotite¹⁷ at a depth of 1,500 km (corresponding to a pressure of 64 GPa). Extrapolation to 135 GPa yields a melting temperature of MORB of about 4,000 K at the core–mantle bound-

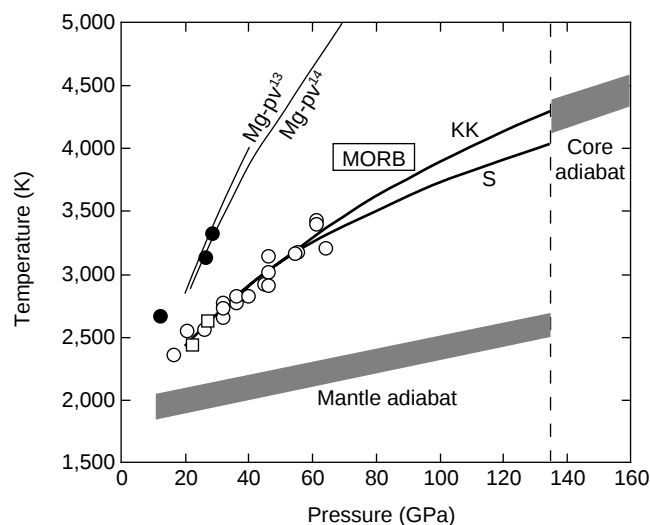


Figure 3 Melting curve of MORB extrapolated to the core–mantle boundary using the melting relationships of Simon²² (S) and Kraut and Kennedy²³ (KK). Open and solid circles represent melting temperatures of MORB and MgSiO₃, respectively, determined in a laser-heated diamond cell. Open squares represent melting temperatures of MORB, determined in the multi-anvil apparatus. Melting curves of St¹⁴, Ca–pv^{14,15} and Mg–pv^{13,14} are substantially higher than that of MORB. The melting curves of Mg–pv from refs 13 and 14 (labelled as Mg–pv¹³ and Mg–pv¹⁴, respectively) are plotted for comparison. The mantle adiabats are from Boehler²³.

ary, which is substantially lower than that of single-phase MgSiO₃–perovskite (Fig. 3). If the temperature locally reaches 4,000 K in the D' region, which may be a graveyard for subducted lithosphere, the former basaltic crust could partially melt. This may provide an explanation for the recent seismic observations of the seismic anisotropy^{18,19} and anomalously slow P-wave velocities^{20,21} at the base of the mantle. Under such a scenario, the temperature of the outer core must be higher than the 4,000 K required for melting of MORB perovskite (Fig. 3). The temperature difference over the thermal boundary between core and mantle may reach 1,500 K, and hot mantle plumes, including partially molten slab materials, are likely to arise from this depth. □

Methods

We have determined the phase relations and the melting temperatures of MORB over a wide pressure range by using two complementary high-pressure techniques, the multi-anvil apparatus and the diamond-anvil cell. The multi-anvil experiments provide detailed chemical composition information of the individual minerals in the high-pressure assemblages through the electron microprobe and structure data through X-ray diffraction. The phase identification was also confirmed by micro-Raman spectroscopy. The temperatures in the multi-anvil experiments were measured with a W₅Re–W₂₆Re thermocouple, whereas pressures were determined based on fixed pressure calibration points, including the ilmenite–perovskite transition in MgSiO₃ (ref. 3). Pressure calibration at high temperature was based on the Al₂O₃ solubility in the MgSiO₃–perovskite structure at pressures above 23 GPa (ref. 6). Combining accurate temperature measurements with a self-consistent pressure calibration, the present multi-anvil experiments provide a reliable determination of the dP/dT slope of the majorite–perovskite transformation in MORB and relatively accurate measurements of melting temperature up to 27 GPa, the maximum attainable pressure in the multi-anvil apparatus.

To extend the melting-temperature measurements for MORB to pressures higher than 27 GPa, we conducted experiments in a laser-heated diamond-anvil cell. In the present diamond cell experiments, a double-sided heating system with a multimode Nd:YAG laser was used to minimize both axial and radial temperature gradients in the heated sample¹¹. The MORB glass sample was sandwiched between two Re foils with Al₂O₃ layers, as thermal insulators, on the top and bottom of the sandwich assemblage. The interface between the

Re and Al_2O_3 were heated with laser beams from both sides. Acting like planar heat sources, the two 'hot plates' eliminate the axial temperature gradient in the sample between the plates. Temperature variation is less than 3% within roughly 30 μm diameter at 2,500 K. Before the melting experiments, the sample was scanned with a laser beam and heated to about 2,000 K to reduce the pressure gradient and to produce a high-pressure solid-phase assemblage. For stable and smooth temperature control, temperatures were increased by adjusting an aperture placed near the beam exit, stepwise, instead of by adjusting power. Each step corresponds to a 50–100 K increase. A 30- μm spot was homogeneously heated by opening the aperture (increasing the step). At the onset of melting, temperature remains constant or drops slightly with the step increment, and then drastically increases (>400 K) within one step. To ensure the reliability of the melting criteria used in this study, we conducted melting experiments at pressures (16–27 GPa) overlapped by the multi-anvil apparatus and the diamond-anvil cell, using the same starting material, and obtained consistent melting temperatures (Fig. 3). We also used the same melting criteria to determine the melting temperature of MgSiO_3 -perovskite previously studied by other investigators, and our results agree with these recent determinations^{13,14} (Fig. 3). The temperature runaway phenomena near the onset of melting observed in simple and complex samples were probably a result of the latent heat of melting, followed by melt migrating away from the heated spot because of the large thermal pressure and, finally, the Re foils would have been heated without sample in between. No chemical reaction between Re and sample was observed in the multi-anvil experiments on a scale of 1 μm . The melting temperatures reported here are the last temperatures before melting sets in. Pressures were measured using a ruby-fluorescence technique after each measurement of melting temperature.

Received 22 June; accepted 12 October 1998.

- Irifune, T. & Ringwood, A. E. Phase transformations in subducted oceanic crust and buoyancy relationships at depths of 600–800 km in the mantle. *Earth Planet. Sci. Lett.* **117**, 101–110 (1993).
- Ringwood, A. E. Role of the transition zone and 660 km discontinuity in mantle dynamics. *Phys. Earth Planet. Inter.* **86**, 5–24 (1994).
- Bertka, C. M. & Fei, Y. Mineralogy of Matian interior up to core-mantle boundary pressures. *J. Geophys. Res.* **102**, 5251–5264 (1997).
- Irifune, T., Ringwood, A. E. & Hiberson, W. O. Subduction of continental crust and terrigenous and pelagic sediments: an experimental study. *Earth Planet. Sci. Lett.* **126**, 351–368 (1994).
- Irifune, T., Koizumi, T. & Ando, J. An experimental study of the garnet-perovskite transformation in the system MgSiO_3 - $\text{Mg}_2\text{Al}_2\text{SiO}_5$. *Phys. Earth Planet. Inter.* **96**, 147–157 (1996).
- Hirose, K. & Fei, Y. Majorite-perovskite transformation on the join MgSiO_3 - $\text{Mg}_2\text{Al}_2\text{SiO}_5$ and its role in the dynamics at the 660-km discontinuity. *Phys. Earth Planet. Inter.* (submitted).
- Kesson, S. E., Fitz Gerald, J. D. & Shelley, J. M. Mineralogy and dynamics of a pyrolytic lower mantle. *Nature* **393**, 252–255 (1998).
- Faust, J. & Knittle, E. The stability and equation of state of majoritic garnet synthesised from natural basalt at mantle conditions. *Geophys. Res. Lett.* **23**, 3377–3380 (1996).
- Kesson, S. E., Fitz Gerald, J. D. & Shelley, J. M. G. Mineral chemistry and density of subducted basaltic crust at lower-mantle pressures. *Nature* **372**, 767–769 (1994).
- Irifune, T. & Ringwood, A. E. Phase transformation in a harzburgite composition to 26 GPa: implications for dynamical behaviour of the subducting slab. *Earth Planet. Sci. Lett.* **86**, 365–376 (1987).
- Shen, G., Mao, H. & Hemley, R. J. Laser-heated diamond anvil cell technique: double-sided heating with multimode Nd:YAG laser. *Proc. ISAM '96* 149–152 (1996).
- Yasuda, A., Fujii, T. & Kurita, K. Melting phase relations of an anhydrous mid-ocean ridge basalt from 3 to 20 GPa: implications for the behavior of subducted oceanic crust in the mantle. *J. Geophys. Res.* **99**, 9401–9414 (1994).
- Zerr, A. & Boehler, R. Melting of (Mg,Fe)SiO₃-perovskite to 625 kilobars: indication of a high melting temperature in the lower mantle. *Science* **262**, 553–555 (1993).
- Shen, G. & Lazor, P. Measurement of melting temperatures of some minerals under lower mantle pressures. *J. Geophys. Res.* **100**, 17699–17713 (1995).
- Zerr, A., Serghiou, G. & Boehler, R. Melting of CaSiO_3 perovskite to 430 kbar and first in-situ measurements of lower mantle eutectic temperatures. *Geophys. Res. Lett.* **24**, 909–912 (1997).
- Presnall, D. C. & Gasparik, T. Melting of enstatite (MgSiO_3) from 10 to 16.5 GPa and the forsterite (Mg_2SiO_4)-majorite (MgSiO_3) eutectic at 16.5 GPa: implications for the origin of the mantle. *J. Geophys. Res.* **95**, 15771–15777 (1990).
- Zerr, A., Diegeler, A. & Boehler, R. Solidus of Earth's deep mantle. *Science* **281**, 243–246 (1998).
- Kendall, J. M. & Silver, P. Constraints from seismic anisotropy in the lowermost mantle. *Nature* **381**, 409–412 (1996).
- Kendall, J. M. & Silver, P. Investigating causes of D'' anisotropy. *AGU Monogr.* (in the press).
- Williams, Q. & Garnero, E. Seismic evidence for partial melt at the base of Earth's mantle. *Science* **273**, 1528–1530 (1996).
- Revenaugh, J. S. & Meyer, R. Seismic evidence of partial melt within a possibly ubiquitous low velocity layer at the base of the mantle. *Science* **277**, 670–673 (1997).
- Simon, F. E. & Glatzel, G. Fusion-pressure curve. *Z. Anorg. Allg. Chem.* **178**, 309 (1929).
- Kraut, E. A. & Kennedy, G. C. New melting law at high pressures. *Phys. Rev. Lett.* **16**, 608–609 (1966).
- Boehler, R. Melting temperature of the Earth's mantle and core: Earth's thermal structure. *Annu. Rev. Earth Planet. Sci.* **15**–40 (1996).
- Mao, H. K., Chen, L. C., Hemley, R. J., Jephcoat, A. P. & Wu, Y. Stability and equation of state of CaSiO_3 -perovskite to 134 GPa. *J. Geophys. Res.* **94**, 17889–17894 (1989).

Acknowledgements. We thank W. Minarik, J. Konzett, G. Shen and S. Shieh for discussions. This research was supported by the NSF, the Carnegie Institution of Washington, and Japan Society for the Promotion of Science.

Correspondence and requests for materials should be addressed to Y.F. (e-mail: fei@gl.ciw.edu).

Migration of plutonium in ground water at the Nevada Test Site

A. B. Kersting*, D. W. Efurdt†, D. L. Finnegan†, D. J. Rokop†, D. K. Smith* & J. L. Thompson†

* Isotope Sciences Division, PO Box 808, L-231, Lawrence Livermore National Laboratory, Livermore, California 94550, USA

† Chemical Science and Technology Division, MS J514, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

Mobile colloids—suspended particles in the submicrometre size range—are known to occur naturally in ground water^{1,2} and have the potential to enhance transport of non-soluble contaminants through sorption³. The possible implications of this transport mechanism are of particular concern in the context of radionuclide transport. Significant quantities of the element plutonium have been introduced into the environment as a result of nuclear weapons testing and production, and nuclear power-plant accidents. Moreover, many countries anticipate storing nuclear waste underground. It has been argued that plutonium introduced into the subsurface environment is relatively immobile owing to its low solubility in ground water⁴ and strong sorption onto rocks⁵. Nonetheless, colloid-facilitated transport of radionuclides has been implicated in field observations^{6,7}, but unequivocal evidence of subsurface transport is lacking^{3,8,9}. Moreover, colloid filtration models predict transport over a limited distance resulting in a discrepancy between observed and modelled behaviour³. Here we report that the radionuclides observed in groundwater samples from aquifers at the Nevada Test Site, where hundreds of underground nuclear tests were conducted, are associated with the colloidal fraction of the ground water. The ²⁴⁰Pu/²³⁹Pu isotope ratio of the samples establishes that an underground nuclear test 1.3 km north of the sample site is the origin of the plutonium. We argue that colloidal groundwater migration must have played an important role in transporting the plutonium. Models that either predict limited transport or do not allow for colloid-facilitated transport may thus significantly underestimate the extent of radionuclide migration.

The Nevada Test Site (NTS) was the location of 828 underground nuclear tests conducted by the United States between 1956 and 1992¹⁰ (Fig. 1a). As a result, the NTS contains a large inventory (>10⁸ Ci) of radioactive material deposited in the subsurface and thus provides a unique opportunity for studying the transport of radionuclide contaminants. During an underground nuclear test, temperatures exceed 10⁶ K locally and ~70 tonnes of rock are vaporized and another 700 tonnes of rock are melted for every kiloton (kt) of explosive yield¹¹. Most refractory radionuclide species (for example, actinides (except U), rare earths, and alkaline earths) are incorporated into the melt glass that coalesces at the bottom of the cavity. The volatile species (for example, alkali metals, U, Sb, I, Ru and gases—Ar, Kr, Xe) are more broadly distributed in the cavity and overlying rubble chimney created directly above the cavity¹².

Samples were collected in the northwestern section of the NTS (Fig. 1a) where thick sequences of ash-flow tuffs and rhyolitic lava flows dominate the geology^{13,14}. Hydrological gradients in the study area suggest a southwestward and southward flow of ground water with estimated flow velocities ranging from 1 to 80 m yr⁻¹ (ref. 13). The ER-20-5 well cluster is located 280 m southwest of the Tybo test site and 1.3 km south of the Benham test site (Fig. 1b). The ER-20-5 wells, no. 1 and no. 3, were screened at a depth of 701–784 m and 1,046–1,183 m, respectively.

Groundwater samples were pumped into 200-l drums on three separate occasions over a 16-month period. Unfiltered groundwater

samples were analysed for tritium (^3H), γ -ray-emitting radionuclides and Pu isotopes. Additional 200-l groundwater samples from well no. 1 were collected on the second and third sampling campaigns and filtered in series using 1,000-nm, 50-nm and ~ 7 -nm (100,000 nominal molecular mass) filter sizes. The particulate material ($>1,000$ nm), two colloidal fractions (1,000–50 nm, and 50–7 nm) and the ultrafiltrate or dissolved fraction (< 7 nm) were separately analysed for ^3H , γ -ray-emitting radionuclides and Pu isotopes¹⁵.

Surface soil samples were collected in the vicinity of the Benham and Tybo nuclear test sites and their $^{240}\text{Pu}/^{239}\text{Pu}$ isotope ratios measured. Archived melt glass material collected from the cavity region immediately after the detonation of the Benham and Tybo tests was re-analysed for its $^{240}\text{Pu}/^{239}\text{Pu}$ isotope composition and compared to data previously obtained during the US nuclear test programme. Aliquots from the melt glass were analysed at Los Alamos National Laboratory as well as Lawrence Livermore National

Laboratory to provide external laboratory comparison. Additional experimental details from this study are given elsewhere^{16,17}.

Tritium and low specific activities of cobalt (Co), caesium (Cs), europium (Eu) and Pu radioisotopes were detected in the unfiltered ground water (Fig. 2). The levels of radioactivity measured from ground water collected from well no. 3 are significantly lower (1,000–20 times lower) than those measured in ground water from well no. 1. Eu isotopes were not detected in ground water from well no. 3 but may be present below the level of detection ($\sim 10^{-3} \text{ Bq l}^{-1}$). Figure 2c illustrates that the same radionuclides and similar levels of radioactivity were detected in the ground water samples obtained from well no. 1 during the three different sampling periods. Analogous results were obtained for ground water samples from well no. 3 (data not shown).

Filtering removed most of the radionuclides from solution, but had essentially no effect on the ^3H concentration in the filtrates. We found that $>99\%$ of the Eu and Pu isotopes, $\sim 91\%$ of the Co, and

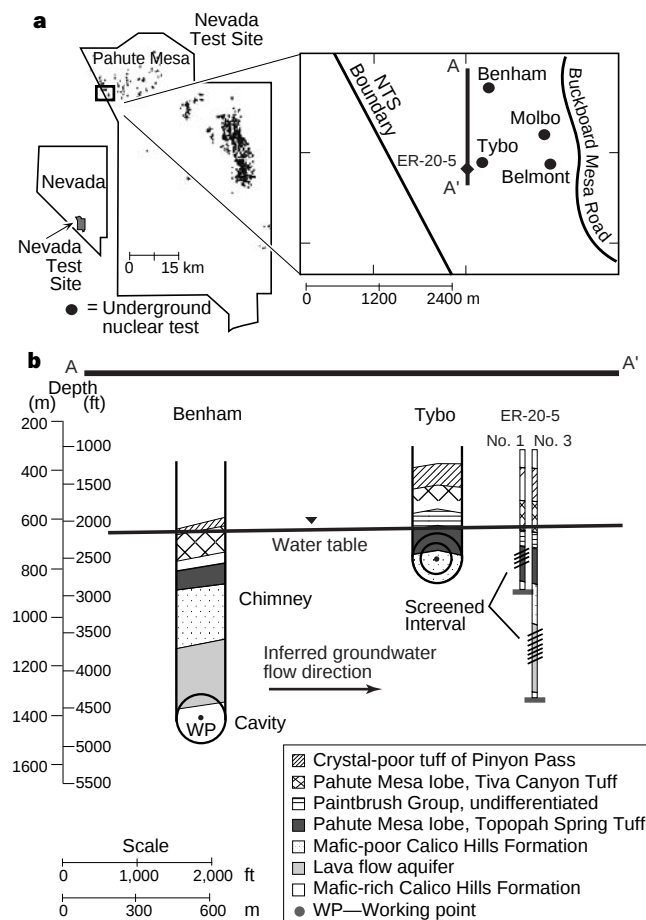


Figure 1 Location and geology of the study area. **a**, Map of the Nevada Test Site, showing the locations of all detonated underground nuclear tests. An enlarged map of the field area in Pahute Mesa is also shown, giving the location of the well cluster ER-20-5 and all other nearby underground nuclear tests. Molbo (1982) and Belmont (1986) have an announced yield between 20 and 150 kt. **b**, A–A' is a north-south cross-section projecting the Benham and Tybo nuclear tests relative to the ER-20-5 well cluster. Well no. 3 is located ~ 30 m south of well no. 1. The Tybo test was detonated in moderately welded tuff on 14 May 1975 at a depth of 765 m and had an announced yield between 200 and 1,000 kt. Benham was detonated in zholitized bedded tuff on 19 December 1968 at a depth of 1,402 m with a nuclear yield of 1,150 kt. The working point (WP) denotes the location of the nuclear device before detonation. The radius of the cavity is a function of the nuclear yield, density of the rock type and depth of burial (distance from ground surface to WP)³¹. Benham has a calculated cavity radius of 98 m, and Tybo between 62 and 105 m based on the unclassified range in yield.

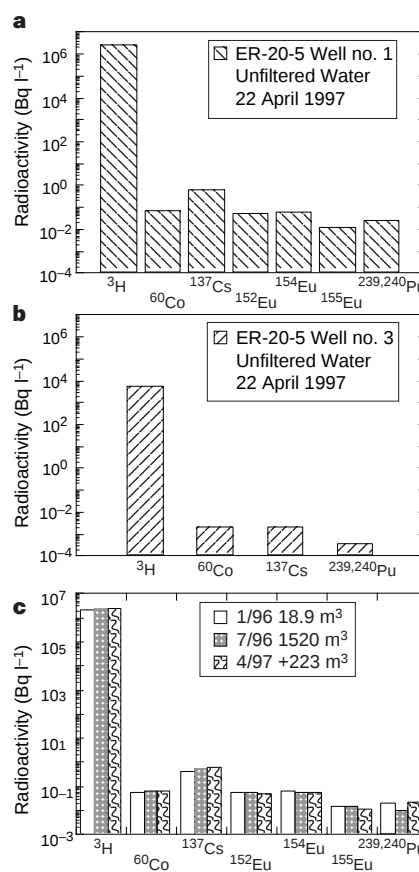


Figure 2 Comparison of radioactivity detected in unfiltered groundwater samples from ER-20-5 well cluster. Groundwater chemistry of ER-20-5 no. 1 in mg l^{-1} is as follows: $[\text{Cl}^-] = 3.3$, $[\text{SO}_4^{2-}] = 35.0$, $[\text{Na}^+] = 70.0$, $[\text{K}^+] = 3.1$, $[\text{Ca}^{2+}] = 3.2$, $[\text{Mg}^{2+}] = 0.1$, pH 8.4. For the purposes of comparison, all data were decay-corrected to 22 April 1997, the time of the third sampling. **a**, Concentration of radioactivity measured in unfiltered ground water from the shallower aquifer, pumped from well no. 1. **b**, Concentration of radioactivity measured in unfiltered ground water from the deeper aquifer, pumped from well no. 3. **c**, Comparison of the radioactivity detected in the three groundwater samples collected from well no. 1 over the duration of the field experiment. The first sample was collected after 18.9 m^3 (5,000 gallons) were pumped; the second after $1.52 \times 10^3 \text{ m}^3$ (401,000 gallons). The pumps were shut down, and restarted the next spring. An additional $2.23 \times 10^2 \text{ m}^3$ (59,000 gallons) of ground water was pumped before the third sample was collected. A total of $1.76 \times 10^3 \text{ m}^3$ were pumped from well no. 1 and $2.19 \times 10^3 \text{ m}^3$ from well no. 3. Pumping rate is $0.030 \text{ m}^3 \text{ min}^{-1}$.

95% of the Cs in the ground water from well no. 1 were associated with the colloidal and particulate fractions (Fig. 3a).

To determine the source of the radionuclides, and hence the distance they have been transported, we use the $^{240}\text{Pu}/^{239}\text{Pu}$ isotope ratio to 'fingerprint' the source of the observed Pu in the ground water. Figure 3b shows the normalized $^{240}\text{Pu}/^{239}\text{Pu}$ isotope ratios measured in this study. The $^{240}\text{Pu}/^{239}\text{Pu}$ isotope ratio of the unfiltered ground water from well no. 1 matches those of the unfiltered ground water from well no. 3 and the colloidal and particulate fraction from well no. 1. In addition, it uniquely matches the Benham nuclear test and no other. (Each nuclear detonation in the study area is characterized by a unique $^{240}\text{Pu}/^{239}\text{Pu}$ isotope ratio.) The $^{240}\text{Pu}/^{239}\text{Pu}$ isotope ratio of the ground water is distinctly different from that of the surface soil samples, eliminating surface contamination as a possible source for the Pu detected in the ground water.

The particulates and two different colloidal size fractions were analysed by X-ray diffraction (XRD) and scanning electron micro-

scopy (SEM). The material was composed of clays (illite and smectite), zeolites (mordenite and clinoptilolite/heulandite), and cristobalite. The same mineral assemblage was detected in all three size fractions ($>1\text{ }\mu\text{m}$, $1,000\text{--}50\text{ nm}$, and $50\text{--}7\text{ nm}$). Figure 4 shows SEM images of two distinct morphologies observed; the flat platy minerals and the rod shaped minerals are likely to be clays (illite) and zeolite (mordenite), respectively.

Clays and zeolites are common secondary minerals in altered rhyolitic tuff and smectite, clinoptilolite and mordenite have been specifically identified in the rocks on Pahute mesa¹³ and elsewhere in volcanic tuffs at the NTS¹⁸. Smectite has also been identified as an alteration mineral formed during dissolution of nuclear waste glass¹⁹. The colloidal minerals identified in ER-20-5 ground water from well no. 1 are consistent with the secondary minerals observed in ground water collected from volcanic tuff aquifers located in the southwestern part of the NTS^{20,21}.

Strontium, Cs, Co and actinides (including Pu) have been shown to strongly sorb to clays and zeolites as a result of their large cation-exchange capacity^{5,22,23}. For example, Pu sorption experiments involving clinoptilolite with grain sizes between $75\text{ and }500\text{ }\mu\text{m}$ and NTS water with a pH of 7 yielded distribution coefficients of $K_d > 500\text{ ml g}^{-1}$ (ref. 5). The zeolite particles and colloids observed in the present study fall within the micrometre and submicrometre size range. Their reduced size will be associated with larger surface areas and is therefore likely to result in higher K_d values. In fact, recent experiments on the sorption of Pu to montmorillonite suggest that this process may be irreversible on the timescale of the laboratory experiments²⁴ (~ 1 year).

The data obtained in this study suggest that Pu and other radionuclides are transported as colloidal material. Although Pu has been shown experimentally to strongly sorb to clays and zeolites, Pu can also exist as an intrinsic colloid, composed of Pu oxide²⁵. Both types of colloids have the capacity to be transported by ground water. From this study, we cannot distinguish the colloidal form of the Pu and further study is needed. We suggest that Co, Cs and Eu are sorbed onto the colloidal sized clays and zeolites in the ground water as they do not form intrinsic colloids.

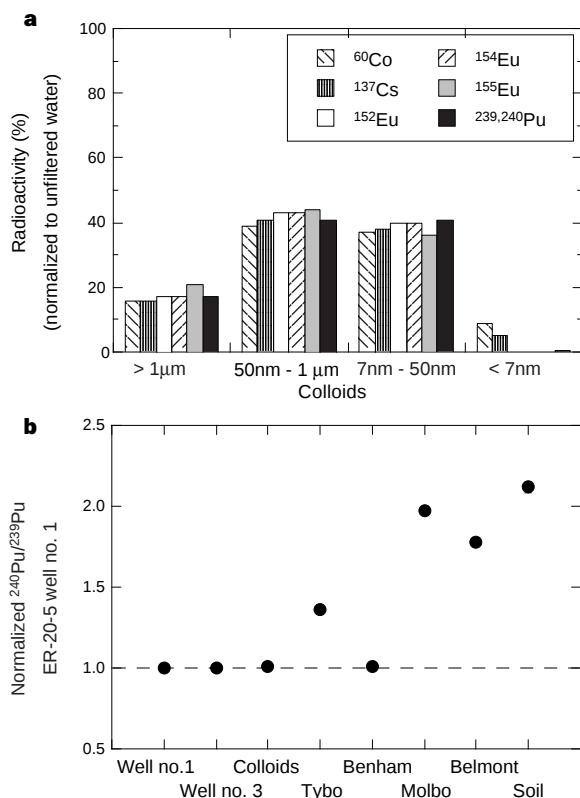


Figure 3 Radioactivity of the colloidal minerals and comparison of Pu isotope ratios from ER-20-5 ground waters to other nuclear tests. **a**, Comparison of the radioactivity of the colloids collected on the different filter sizes and the ultrafiltrate fraction. Data are normalized to the total radioactivity measured in the unfiltered water. The ultrafiltrate is the water that passed through all the filters ($<7\text{ nm}$). Filtration occurred in series. A tangential filtration system was used for the $100,000$ nominal molecular mass ($\sim 7\text{ nm}$) size filters. **b**, Comparison of the $^{240}\text{Pu}/^{239}\text{Pu}$ isotope ratios of different samples in this study normalized to the radioactivity measured in ER-20-5 no. 1. Precision is $\pm 1.5\%$. The errors plotted are smaller than the symbols used. Results of the Pu isotopic analyses of the archived melt glass material collected from the cavity region immediately after the detonation of Benham and Tybo from both laboratories agree to within 1.5% and also match the original values measured immediately following the nuclear tests. Total laboratory procedural blanks had $<2\text{ pg Pu}$, significantly below the concentrations analysed and did not contribute to the isotope ratio measured. The concentrations of Pu detected in the soil samples were extremely low ($<4\text{ pg Pu per g sample}$), and the isotopic ratios distinctly different from the ground water. Values plotted are averages: ER-20-5 no. 1, $N = 3$; ER-20-5 no. 3, $N = 2$; Colloids, $N = 4$; Tybo, $N = 6$; Benham, $N = 12$; Molbo, $N = 7$; Belmont, $N = 4$; and soil, $N = 2$. Here N is the number of samples averaged.

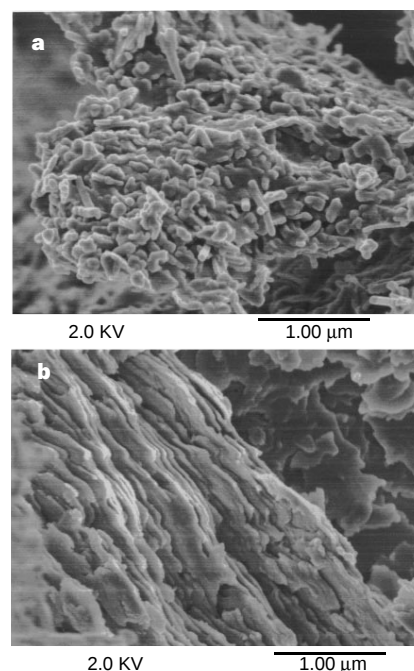


Figure 4 High-resolution SEM images of the colloids in ER-20-5 no. 1. **a**, The tabular, lath-shaped morphology of the zeolite, mordenite; **b**, the platy appearance of the clay, illite. The two distinct morphologies were observed in all three size fractions ($>1\text{ }\mu\text{m}$, $1,000\text{--}50\text{ nm}$, and $50\text{--}7\text{ nm}$).

The maximum measured concentration of Pu at the ER-20-5 site is $\sim 10^{-14}$ M. This value is lower than the solubility limits of $\sim 10^{-8}$ M that have been experimentally determined for the Pu (v) species likely to be present in NTS ground water⁴. The calculated solubility limits (10^{-12} – 10^{-17} M), obtained by assuming that Pu(IV) is in thermodynamic equilibrium^{26,27}, bracket the maximum measured Pu concentration. It thus seems that the Pu concentrations in the ground water at ER-20-5 were too low to lead to the precipitation of a solid Pu phase. Our results indicate that <1% of the observed Pu is in the dissolved fraction of the ground water. This finding and the previously reported results of Pu sorption experiments^{4,5} are most consistent with Pu migrating as colloidal material and not as a dissolved phase. The concentration of Pu measured is small, and represents only a small fraction of the total Pu associated with the Benham nuclear test.

Based on 40 years of re-drilling underground nuclear test cavities and collecting melt glass samples for test diagnostics, it has been observed that the majority (~98%) of the refractory radionuclides (such as Pu) are incorporated into the melt glass that forms at the bottom of the test cavity^{12,28}. In field studies where ground water can be eliminated as a possible transport mechanism, radionuclides were detected at a maximum of a few hundred metres from the original detonation point, and were attributed to gas movement through fractures, or fracture injection of vaporized material at detonation time^{29,30}. The possibility that Pu from the Benham test site was blasted and deposited >1.3 km away, in two distinct aquifers separated by 300 m vertically and 30 m horizontally, seems highly unlikely. However, some fraction of the Pu may have been initially injected through fractures a few hundred metres and subsequently transported by ground water.

Molbo, Belmont and Tybo nuclear tests were all detonated after Benham. Although shock waves resulting from underground nuclear blasts can induce radial fractures out to a maximum distance of a few hundred metres (ref. 30), it is unlikely that these detonations blasted material from Benham to ER-20-5, as they were all smaller detonations and by inference shallower. In addition, Pu from these subsequent three tests was not detected in the ground water at ER-20-5.

The high pumping rates ($0.03 \text{ m}^3 \text{ min}^{-1}$) employed may create shear stresses sufficient to generate an increase in the concentration of colloids and thus prevent a quantification of the ambient colloidal load and, by inference, a determination of the minimum or maximum concentration of Pu in the ground water. But the isotope ratio of the Pu measured in ground water from ER-20-5 nevertheless clearly establishes that the radionuclide originates from a specific nuclear event, ~1.3 km to the north. The present work thus demonstrates that Pu is not immobile in the subsurface, but can be transported over significant distances. Pu transport models that only take into account sorption and solubility may therefore underestimate the extent to which this species is able to migrate in ground water. □

Received 24 February; accepted 6 October 1998.

1. Degueldre, C. et al. Colloids in water from a subsurface fracture in granitic rock, Grimsel test site, Switzerland. *Geochim. Cosmochim. Acta* **53**, 603–610 (1989).
2. McDowell-Boyer, L. M. Chemical mobilization of micron sized particles in saturated porous media under steady flow conditions. *Environ. Sci. Technol.* **26**, 586–593 (1992).
3. Ryan, J. N. & Elimelech, M. Colloid mobilization and transport in groundwater. *Colloids Surfaces A: Physicochem. Eng. Aspects* **107**, 1–56 (1996).
4. Nitsche, H. et al. Measured Solubilities and Speciations of Neptunium, Plutonium, and Americium in a Typical Groundwater (J-13) from the Yucca Mountain Region Milestone Report 3010-WBS 1.2.3.4.1.3.1 (Rep. LA-12562-MS, Los Alamos National Laboratory, 1993).
5. Triay, I. R. et al. Radionuclide Sorption in Yucca Mountain Tuffs with J-13 Well Water: Neptunium, Uranium, and Plutonium (Rep. LA-12956-MS, Los Alamos National Laboratory, 1996).
6. Buddemeier, R. W. & Hunt, J. R. Transport of colloidal contaminants in groundwater: radionuclide migration at the Nevada Test Site. *Appl. Geochem.* **3**, 535–548 (1988).
7. Penrose, W. R., Polzer, W. L., Essington, E. H., Nelson, D. M. & Orlandini, K. A. Mobility of plutonium and americium through a shallow aquifer in a semiarid region. *Environ. Sci. Technol.* **24**, 228–234 (1990).
8. McCarthy, J. F. & Degueldre, C. in *Environmental Particles 2* (eds Buffle, J. & van Leeuwen, H. P.) 247–315 (Lewis, Ann Arbor, 1993).
9. Marty, R. C., Bennett, D. & Thullen, P. Mechanism of plutonium transport in a shallow aquifer in

- Mortandad canyon, Los Alamos National Laboratory, New Mexico. *Environ. Sci. Technol.* **31**, 2020–2027 (1997).
10. *United States Nuclear Tests—July 1945–September 1992* (DOE/NV-209 (Rev. 14), US Department of Energy/Nevada Field Office, 1994).
11. Smith, D. K. Characterization of nuclear explosive melt debris. *Radiochim. Acta* **69**, 157–167 (1995).
12. Borg, I. Y., Stone, R., Levy, H. B. & Ramsdottir, L. D. *Information Pertinent to the Migration of Radionuclides in Ground Water at the Nevada Test Site* (Rep. UCRL-52078, Lawrence Livermore National Laboratory, 1976).
13. Blankennagel, R. K. & Weir, J. E. J. *Geohydrology of the eastern part of Pahute Mesa, Nevada Test Site, Nye County, Nevada Geol. Surv. Prof. Pap. 712-B* (1973).
14. Lacznak, R. J., Cole, J. C., Sawyer, D. A. & Trudeau, D. A. *Summary of Hydrogeologic Controls on Ground-Water Flow at the Nevada Test Site, Nye County, NV* (Rep. 96-4109, US Geological Survey, 1996).
15. Perrin, R. E., Knobeloch, G. W., Armijo, V. M. & Efurud, D. W. Isotopic analysis of nanogram quantities of plutonium by using a SID ionization source. *Int. J. Mass Spectrom. Ion Phys.* **64**, 17–24 (1985).
16. Thompson, J. L. *Laboratory and Field Studies Related to radionuclide Migration at the Nevada Test Site October 1, 1996–September 30, 1997* (Rep. LA-13419-PR Ed., Los Alamos National Laboratory, 1998).
17. Smith, D. K. et al. *Hydrologic Resources Management Program and Underground Test Area Operable Unit FY 1997 Progress Report* (Rep. UCRL-ID-130792 Ed., Lawrence Livermore National Laboratory, 1998).
18. Broxton, D. E., Bish, D. L. & Warren, R. G. Distribution and chemistry of diagenetic minerals at Yucca Mountain, Nye County, Nevada. *Clays Clay Miner.* **35**, 89–110 (1987).
19. Bates, J. K., Bradley, J. P., Teetsov, A., Bradley, C. R. & Buchholtz ten Brink, M. Colloid formation during waste form reaction: implications for nuclear waste disposal. *Science* **256**, 649–651 (1992).
20. Viani, B. E. & Martin, S. I. *Groundwater Colloid Characterization* (UCRL–Lawrence Livermore National Laboratory, 1996).
21. Kingston, W. L. & Whitbeck, M. *Characterization of Colloids Found in Various Groundwater Environments in Central and Southern Nevada* (Rep. DOE-NV/10384-36, Desert Research Inst., Las Vegas, 1991).
22. Comans, R. N. J. & Hockley, D. E. Kinetics of cesium sorption on illite. *Geochim. Cosmochim. Acta* **56**, 1157–1164 (1992).
23. Torstenfeld, B., Rundberg, R. S. & Mitchell, A. J. Actinide sorption on granites and minerals as a function of pH and colloids/pseudocolloids. *Radiochim. Acta* **44/45**, 111–117 (1988).
24. Triay, I. R., Lu, N., Cotter, C. R. & Kitten, H. D. *Iron Oxide Colloid Facilitated Plutonium Transport in Groundwater* (Am. Chemical Soc., Las Vegas, 1997).
25. Silva, R. J. & Nitsche, H. Actinide environmental chemistry. *Radiochim. Acta* **70/71**, 377–396 (1995).
26. Stout, R. B. & Leider, H. *Preliminary Waste Form Characteristics Report Version 1.0* (Rep. UCRL-ID-108314 Rev. 1, Lawrence Livermore National Laboratory, 1994).
27. Guillaumont, R. & Adloff, J. P. Behavior of environmental plutonium at very low concentrations. *Radiochim. Acta* **58/59**, 53–60 (1992).
28. Levy, H. B. *On Evaluating the Hazards of Groundwater Contamination by Radioactivity from an Underground Nuclear Explosion* (Rep. UCRL-51278, Lawrence Livermore National Laboratory, 1972).
29. Nimz, G. J. & Thompson, J. L. *Underground Radionuclide Migration at the Nevada Test Site* (Rep. DOE/NV-346, US Department of Energy, Nevada Field Office, 1992).
30. Smith, D. K., Nagle, R. J. & Kenneally, J. M. Transport of gaseous fission products adjacent to an underground nuclear test cavity. *Radiochim. Acta* **73**, 177–183 (1996).
31. Hudson, B. C., Jones, E. M., Keller, C. E. & Smith, C. W. *Cavity Radius Uncertainties* (Rep. LA-9211-C, Los Alamos National Laboratory, Monterey, 1981).

Acknowledgements. We thank B. A. Martinez, F. R. Roensch, J. W. Chamberlin and G. P. Russ for help with sample analysis, and I. R. Triay, B. E. Viani, R. W. Loughheed, H. F. Shaw, F. J. Ryerson, J. F. Wild and G. B. Hudson for discussions. This work was funded partly by the Underground Test Area Project sponsored by US Department of Energy, Nevada Operations Office. Work was performed under the auspices of the US Department of Energy by Los Alamos National Laboratory and Lawrence Livermore National Laboratory.

Correspondence and requests for materials should be addressed to A.B.K. (e-mail: Kersting@llnl.gov).

Patterns of recruitment and abundance of corals along the Great Barrier Reef

T. P. Hughes, A. H. Baird, E. A. Dinsdale, N. A. Moltschaniwskyj*, M. S. Pratchett, J. E. Tanner & B. L. Willis

Department of Marine Biology, James Cook University, Townsville, Queensland 4811, Australia

Different physical and biological processes prevail at different scales^{1–4}. As a consequence, small-scale experiments or local observations provide limited insights into regional or global phenomena^{5–8}. One solution is to incorporate spatial scale explicitly into the experimental and sampling design of field studies, to provide a broader, landscape view of ecology^{1–8}. Here we examine spatial patterns in corals on the Great Barrier Reef, across a spectrum of scales ranging from metres to more than 1,700 km. Our study is unusual because we explore large-scale patterns of a

* Present address: Department of Aquaculture, University of Tasmania, Launceston, Tasmania 7250, Australia.

process (recruitment by juveniles) as well as patterns of adult abundance, revealing the relationship between the two. We show that coral-reef assemblages that are similar in terms of abundance may nonetheless show profound differences in dynamics and turnover, with major implications for their ecology, evolution and management.

The continuous expanse of Australia's Great Barrier Reef (GBR) provides a unique opportunity for exploring the effects of scale dependency in marine systems. The GBR Marine Park is the largest underwater reserve in the world, stretching over 2,000 km and comprising some 2,500 individual reefs separated by distances ranging from a few hundred metres to tens of kilometres. Understanding patterns of recruitment among reefs by larvae is crucial for managing this system. Previous studies of coral recruitment on the GBR⁹ have been limited to adjacent sites or to very few reefs, and have used a variety of techniques that preclude statistical comparisons at larger scales. Here we measured rates of recruitment on 33 reefs that stretch from 10° to 23° S (Fig. 1). We used a hierarchical sampling design¹⁰ to measure spatial variation in recruitment and

adult abundance, at four scales spanning 5–6 orders of magnitude (Fig. 2). This approach provides a powerful framework for quantifying the proportion of the total variation among samples that is attributable to each spatial scale. We distinguish here between two ecologically distinct groups of corals: spawners and brooders. (Most corals on the GBR are spawners, that is, they release sperm and eggs simultaneously^{9,11}. In contrast, brooders release fertilized larvae. The dominant brooders on the GBR are three species of *Acropora* in the subgenus *Isopora*, and eight species of *Pocillopora*, *Seriatopora* and *Stylophora* in the family Pocilloporidae.)

The density of recruits varied at all spatial scales for both spawners and brooders, with distinctive patterns occurring in each group. Recruitment by spawners varied by more than 25-fold among sectors in both years, compared with a more modest five- to eightfold variation for brooders (Fig. 3). In year 1, the two types of recruit were equally abundant at the northern- and southern-most sectors (1 and 6, respectively), whereas spawners were 3–5 times more numerous elsewhere. In year 2, spawning recruits exceeded brooders in all sectors. For spawners, 52–55% (depending

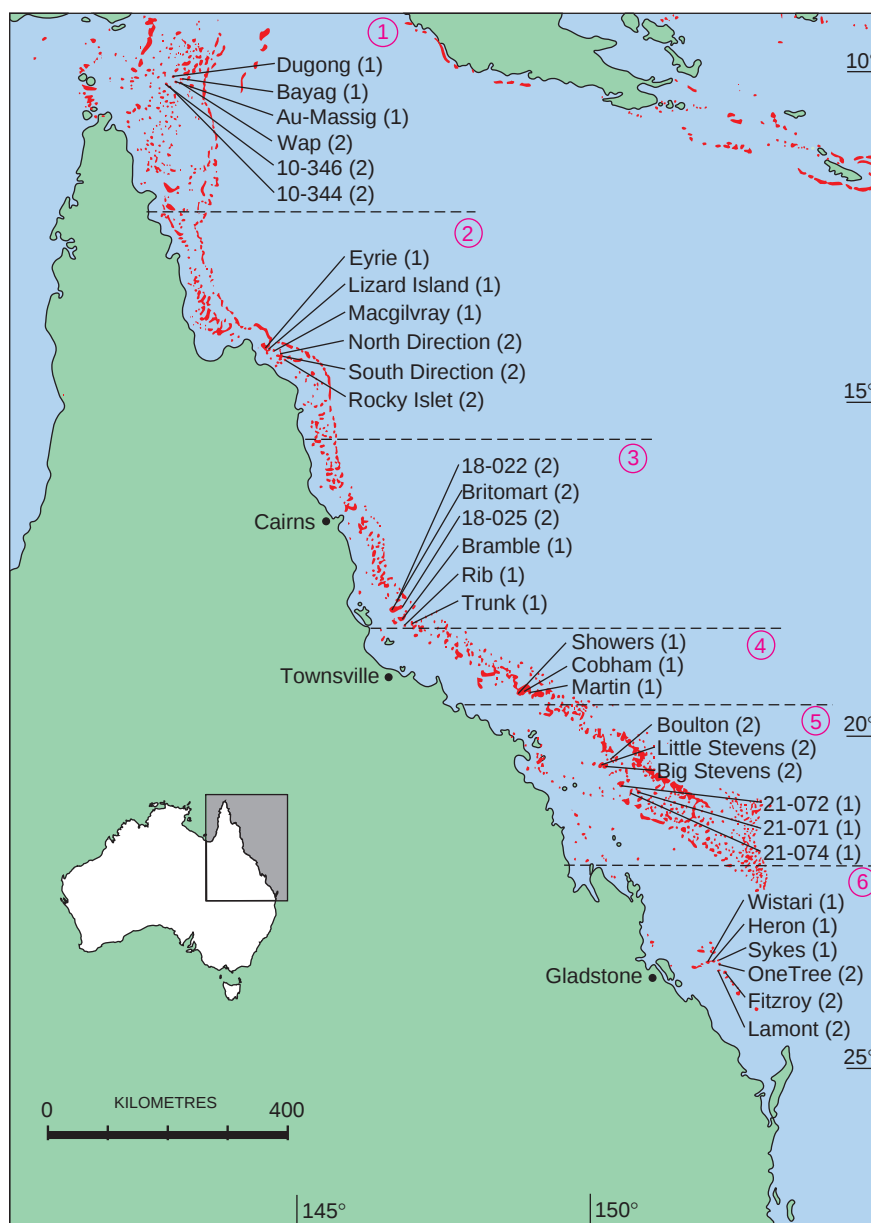


Figure 1 Map of the GBR, indicating the location of 33 reefs at which recruitment and abundance of corals were measured. Eighteen reefs in six sectors were

sampled in 1995/6 (year 1), and a further 15 reefs in 1996/7 (year 2). Sectors of the GBR are numbered 1–6 from north to south.

on year) of the variation in recruitment was attributable to the largest spatial scale (sector), reflecting the marked latitudinal differences in density (Fig. 4). In contrast, variability in recruitment by brooders among sectors accounted for only 25–29% of the total variation. Variability was low at the scale of adjacent reefs for both spawners and brooders (0–6% of the total variation for spawners, and 3–13% for brooders). At smaller scales within reefs, variation increased once more: 21–22% and 25–36% of the total variation occurred among sites on a reef for spawners and brooders, respectively. Finally, variation among replicate panels within sites accounted for 18–26% of the total variation in spawners, compared with 32–36% in brooders. The partitioning of variation among spatial scales was highly consistent over two years for both spawners and brooders (Fig. 4), despite a threefold difference among years in the total amount of recruitment (24.8 ± 1.1 (s.e.) recruits per panel in year 1 compared with 86.9 ± 4.2 recruits in year 2).

Recruitment by spawners varied most among sectors and least among reefs in both years (Fig. 4). This pattern is driven presumably by large-scale hydrodynamic or meteorological phenomena, such as the latitudinal gradient in water temperature along the east coast of Australia. At a much smaller scale, the mean number of recruits varied among sites by more than 200-fold for spawners and 100-fold

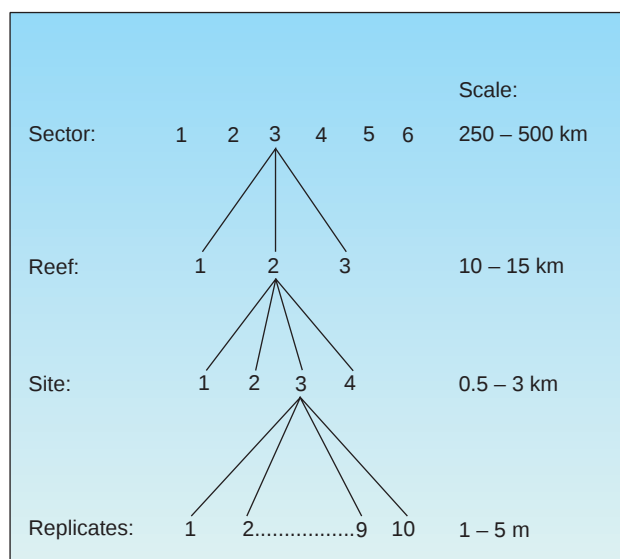


Figure 2 A hierarchical sampling design used to measure spatial variation in the recruitment and abundance of corals in sectors along the GBR, each 250–500 km apart. Within each sector, we selected three mid-shelf reefs at random, separated by ~10–15 km (Fig. 1). On each reef, we chose four sites on the reef crest that were 0.5–3 km apart. Within-site variation at a scale of a few metres was measured among replicates (ten replicate panels for measuring recruitment, and ten 10-m-long transects for measuring adult abundance). Panels were deployed synchronously on mid-shelf reef crests (approximately 1 m below datum), ten days (± 1) before the predicted annual mass-spawning of corals^{9,11}, and retrieved eight weeks later. Panels were unglazed tiles (11 × 11 cm) attached individually to the substratum by a bolt that held them 2–3 cm above the reef surface. Note that the panels were deployed on different reefs in separate years as our objective was to measure the effects of spatial scale rather than the predictability of recruitment at any particular site or reef. A total of 1,135 panels were relocated and recruits on their undersurface were counted. Juvenile corals were identified to family or genus where possible, and classified as spawners or brooders⁹. In year 1, the abundance of adult spawning and brooding corals was measured using ten 10-m-long intercept transects, placed alongside the recruitment panels. All colonies were identified, counted and measured to the nearest centimetre. The number of recruits each year, cover of adults and the number of adults were analysed separately for spawners and brooders using a three-factor nested analysis of variance, with sectors, reefs and sites as random factors. Heterogeneity of variances was removed using $\log(x + 1)$ transformations.

for brooders in all sectors in both years. This level of variation (see also Fig. 5) is unlikely to have been due to spatial variation in mortality after settlement, as the panels were exposed for only 8 weeks and there were few signs of predation or competitive interactions on them (particularly on the undersurfaces of panels where the recruits were counted). We propose therefore that most of the variation in patterns of recruitment (Fig. 4) occurred at the time of settlement, at both large and small scales. Furthermore, the greater variation among reefs and sites for brooders (39% of the total variation in both years compared with 22–27% for spawners) is consistent with their generally shorter larval duration⁹ and lower levels of gene flow (refs 12–14 and D. J. Ayre and T.P.H., unpublished data).

Sectors, reefs, sites and panels that had large numbers of spawners did not necessarily receive large numbers of brooders (Figs 3, 5). Similarly, large-scale patterns of recruitment by reef fishes also differ among species^{15,16}. These disparities have important implications for rates of re-colonization by different taxa after disturbances (for example, as a result of cyclones, mass bleaching or outbreaks of crown-of-thorns starfish), and for the design and management of coral-reef reserves. Clearly, sectors or reefs that are ‘sources’ or ‘sinks’ for one taxonomic group may show the opposite pattern for another, depending on factors such as stock size, larval duration, planktonic mortality, and settlement behaviour.

In contrast to patterns of recruitment, the abundance of adult spawning corals (number of colonies per transect) did not vary significantly among sectors or reefs (Fig. 3), with 91% of the total variation occurring among sites and among replicate transects (Fig. 4). Adult brooding corals varied threefold among sectors, accounting for 29% of the overall variation. As with spawners, the density of adult brooders did not differ significantly among adjacent reefs, and the remainder of the variation occurred at smaller scales. Spawners were more abundant than brooders in all sectors and on all reefs (Fig. 3). Nearly identical patterns were found for adult cover. These results show that the total abundance of adult spawning corals on reef crests along the GBR is relatively homogeneous among reefs and sectors, despite substantial differences in recruitment at large scales (Figs 3, 4). The same pattern occurs at an oceanic scale: corals in the Caribbean Sea can achieve high adult numbers and cover that are comparable to the GBR, yet rates of recruitment there are ~2 orders of magnitude lower (typically < 1 per panel after a few months^{17–19}). The lack of conformity between adult abundance and recruitment implies that rates of post-recruitment survival are higher in geographic regions or sectors of the GBR

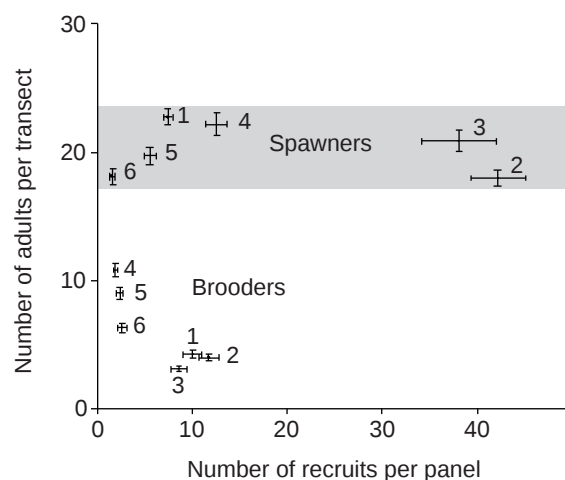


Figure 3 Amounts of recruitment (number per panel) and adult abundance (number of colonies per 10-m transect) of spawning and brooding corals on reef crests in six sectors of the GBR (numbered 1–6 from north to south, see Fig. 1). Error bars indicate one standard error among replicate panels and transects.

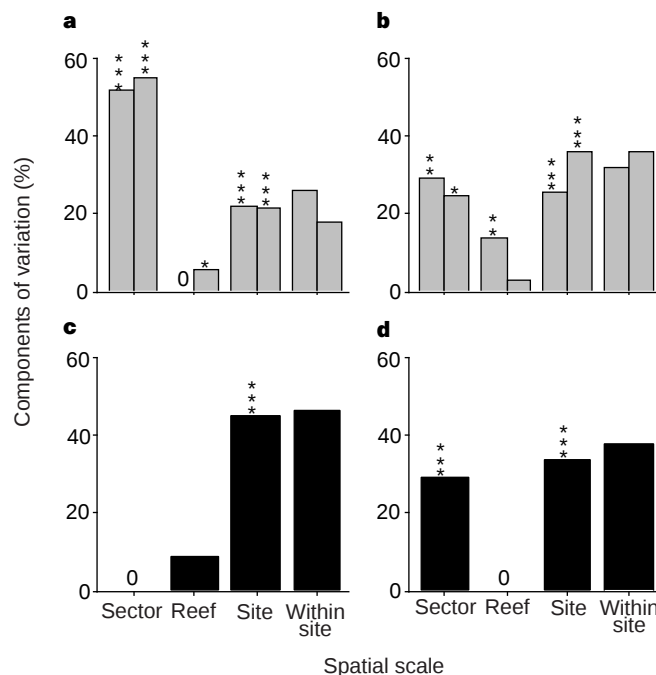


Figure 4 Components of variation at four spatial scales. Variation in numbers of **a**, spawning recruits; **b**, brooding recruits; **c**, spawning adults; and **d**, brooding adults. The overall variation is partitioned among scales, and expressed as a percentage of the total¹⁰. Bars with asterisks indicate the spatial scales where

there is significant variation, that is, where substantial differences occurred among sectors, reefs, and/or sites (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$). Pairs of adjoining bars for recruits in **(a)** and **(b)** are for two separate years, 1995/6 and 1996/7, and different reefs.

that have fewer recruits. Comparative studies of recruitment and survival of reef fishes also support this hypothesis^{20,21}. Alternatively, spatial variability in recruitment could be effectively neutralized if locations that usually have small numbers of recruits occasionally receive a much larger input. Such a scenario should produce dominant year classes within adult populations (providing that post-recruitment mortality is not strongly density dependent). However, coral populations are more typically composed of a multitude of cohorts that have accumulated over many years^{22–24}.

Large-scale variation in recruitment has profound implications for management of coral reefs. For example, the lower recruitment rates onto Caribbean reefs are likely to result in slower rates of recovery from natural or man-made disturbances than on the GBR^{25,26}. The high predictability in the number and cover of adult spawners at the scale of sectors and reefs on the GBR (Fig. 4) does

not mean uniformity of processes (for example, recruitment and mortality) that control abundance. We conclude that management of natural resources based on sound knowledge of process and mechanisms will be far more robust than the traditional approach that has centred on monitoring patterns of abundance. □

Received 26 August; accepted 3 November 1998.

- Kareiva, P. & Andersen, M. in *Lecture Notes in Biomathematics: Community Ecology* (ed. Hastings, A.) 38–45 (Springer, Berlin, 1988).
- Wiens, J. A. Spatial scaling in ecology. *Func. Ecol.* **3**, 385–397 (1989).
- Jackson, J. B. C. Adaptation and diversity of reef corals. *Bioscience* **41**, 475–482 (1991).
- Schneider, D. C. *Quantitative Ecology: Spatial and Temporal Scaling* (Academic, San Diego, 1994).
- Ricklefs, R. E. Community diversity: relative roles of local and regional processes. *Science* **235**, 167–171 (1987).
- Levin, S. The problem of pattern and scale in ecology. *Ecology* **73**, 1943–1967 (1992).
- Ricklefs, R. E. & Schluter, D. (eds) *Species Diversity in Ecological Communities* (Univ. Chicago Press, 1993).
- Eberhardt, L. L. & Thomas, J. M. Designing environmental field studies. *Ecol. Monogr.* **61**, 53–73 (1991).
- Harrison, P. L. & Wallace, C. C. in *Ecosystems of the World* (ed. Dubinsky, Z.) 133–207 (Elsevier, New York, 1992).
- Searle, S. R., Casella, G. & McCulloch, C. E. *Variance Components* (Wiley Interscience, New York, 1992).
- Harrison, P. L. *et al.* Mass-spawning in tropical reef corals. *Science* **223**, 1186–1189 (1984).
- Ayre, D. J. & Dufty, S. L. Restricted gene flow in the viviparous coral *Seriatopora hystrix* on Australia's Great Barrier Reef. *Evolution* **48**, 1183–1201 (1994).
- Benzie, J. A. H., Haskell, A. & Lehman, H. Variation in the genetic composition of coral (*Pocillopora damicornis* and *Acropora palifera*) populations from different reef habitats. *Mar. Biol.* **121**, 731–739 (1995).
- Ayre, D. J., Hughes, T. P. & Standish, R. J. Genetic differentiation, reproductive mode, and gene flow in the brooding coral *Pocillopora damicornis* along the Great Barrier Reef, Australia. *Mar. Ecol. Prog. Ser.* **159**, 175–187 (1997).
- Sale, P. F., Doherty, P. J., Eckert, G. J., Douglas, W. A. & Ferrell, D. J. Large scale spatial and temporal variation in recruitment of fish populations on coral reefs. *Oecologia* **64**, 191–198 (1984).
- Doherty, P. J. & Williams, D. M. B. The replenishment of coral reef fish populations. *Oceanogr. Mar. Biol. Annu. Rev.* **26**, 487–551 (1988).
- Rogers, C. R., Fitz, H. C., Gilnick, M., Beets, J. & Hardin, J. Scleractinian coral recruitment patterns at Salt River submarine canyon, St. Croix, U.S. Virgin Islands. *Coral Reefs* **3**, 69–76 (1984).
- Richmond, R. H. & Hunter, C. L. Reproduction and recruitment of corals: comparisons among the Caribbean, the tropical Pacific, and the Red Sea. *Mar. Ecol. Prog. Ser.* **60**, 185–203 (1990).
- Smith, S. R. Patterns of coral recruitment and post-settlement mortality on Bermuda's reefs: comparisons to Caribbean and Pacific reefs. *Am. Zool.* **32**, 663–673 (1992).
- Sweatman, H. P. A. The influence of adults of some reef fishes on larval recruitment. *Ecol. Monogr.* **55**, 469–485 (1985).
- Caley, M. J. Reef-fish community structure and dynamics: an interaction between local and larger-scale processes? *Mar. Ecol. Prog. Ser.* **129**, 19–29 (1995).
- Connell, J. H. in *Biology and Geology of Coral Reefs*, Vol. 1 (eds Jones, O. A. & Endean, R. E.) 205–245 (Academic, New York, 1973).

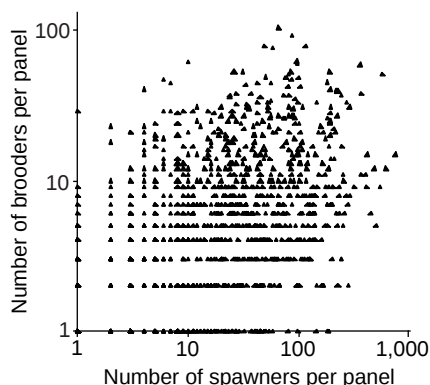


Figure 5 The correlation between recruitment of spawners and brooders (both log(x + 1)-transformed), plotted here at the scale of individual panels ($n = 1,135$ panels for two years combined; $r = 0.27$, $P < 0.001$). The correlation is significant owing to the large sample size. There was only a weak correlation between the two types of recruit at any spatial scale (site, $n = 127$, $r = 0.27$, $P = 0.002$; reef, $n = 33$, $r = 0.42$, $P = 0.014$; sector, $n = 11$, $r = 0.53$, $P = 0.09$).

23. Connell, J. H., Hughes, T. P. & Wallace, C. D. A 30-year study of coral community dynamics: influence of disturbance and recruitment on abundance, at several scales in space and time. *Ecol. Monogr.* **67**, 461–488 (1997).
24. Hughes, T. P. Demographic approaches to community dynamics: a coral reef example. *Ecology* **77**, 2256–2260 (1996).
25. Kojis, B. in *Proc. Colloquium on Global Aspects of Coral Reefs: Health, Hazards, History* (ed. R. N. Ginsburg) P35–P41 (Univ. Miami, Florida, 1993).
26. Connell, J. H. Disturbance and recovery of coral assemblages. *Coral Reefs* **16**(Suppl.), 101–113 (1997).

Acknowledgements. We thank the 41 graduate-student volunteers from James Cook University (JCU) who provided field assistance, and D. Ayre, J. Caley, P. Doherty and L. Smith for comments on the manuscript. Research was funded by grants to T.P.H. from the Australian Research Council. This is contribution no. 162 of the Coral Group at JCU.

Correspondence and requests for materials should be addressed to T.P.H. (e-mail: terry.hughes@jcu.edu.au).

Robust and optimal use of information in stereo vision

John Porrill*, John P. Frisby*, Wendy J. Adams* & David Buckley†

Departments of *Psychology and †Ophthalmology and Orthoptics, University of Sheffield, Sheffield S10 2TN, UK

Differences between the left and right eye's views of the world carry information about three-dimensional scene structure and about the position of the eyes in the head. The contemporary Bayesian approach to perception^{1,2} implies that human performance in using this source of eye-position information can be analysed most usefully by comparison with the performance of a statistically optimal observer. Here we argue that the comparison observer should also be statistically robust, and we find that this requirement leads to qualitatively new behaviours. For example, when presented with a class of stereoscopic stimuli containing inconsistent information about eccentricity of gaze, estimates of this gaze parameter recorded from one robust ideal observer bifurcate at a critical value of stimulus inconsistency. We report an experiment in which human observers also show this phenomenon and we use the experimentally determined critical value to estimate the vertical acuity of the visual system. The Bayesian analysis also provides a highly reliable and biologically plausible algorithm that can recover eye positions even before the classic stereo-correspondence problem is solved, that is, before deciding which features in the left and right images are to be matched.

To be able to specify the observer that is statistically optimal for a perceptual task³, a stochastic model of sensor noise must be available. For mathematical convenience, simple models, such as additive gaussian noise, are generally used. Although such models are adequate for describing some sources of error, most sensory data will also contain a proportion of rogue data items. It is reasonable to assume that perceptual systems have evolved to deal with this problem. We therefore introduce the idea of a robust ideal observer, that is, an observer that has optimal performance in the presence of a suitably specified population of outliers. We take this terminology from statistics, where the population mean is a simple example of a non-robust estimator and the median and mode of robust estimators. Robust and non-robust ideal observers will show quantitatively different levels of performance as the proportion of outliers increases. What is more interesting is that they can also show qualitatively different behaviour under certain experimental conditions, leading to a new type of comparison between ideal observer and human performance.

We have applied the robust ideal observer analysis to the following problem in stereo vision. The disparity in position between left and right retinal images of a point in the field of vision can be decomposed into horizontal and vertical components. The horizontal components H contain information about the three-dimensional structure of the world scene but they cannot be interpreted fully without knowledge of two viewing parameters

that determine the positions of the eyes in the head: g , the gaze angle of the fixation point from the median plane of the head, and d , the distance to the fixation point. It is often more convenient to use the alternative dimensionless parameters⁴ $\alpha = Ig/d$ and $\beta = I/d$, where I is the interocular separation. Because the vertical component of disparity V depends to first order on only these viewing parameters and not on visual scene depths, measurements of V can, in principle, be used to determine the viewing parameters and calibrate the information in H . We use the term V analysis to describe the recovery of viewing parameters from V measurements^{5–10}.

The experimental tool we used to investigate V analysis is the induced effect¹¹. This is produced when one eye's view of a frontoparallel plane is magnified in the vertical dimension. The effect is a perceived rotation of the plane about a vertical axis through an angle that is approximately proportional to the applied magnification. The size of the effect in human observers is consistent with their performing V analysis to derive an incorrect eccentric gaze position and interpreting H information accordingly^{6–10}. A convenient way to measure the induced effect is to have the observer vary a second magnification applied in the horizontal dimension which also produces a perceived slant about a vertical axis, an effect called the geometric effect. The horizontal magnification can be adjusted until the geometric and induced effects cancel and the surface appears frontoparallel.

Human and ideal observer performances were compared on this nulling task. The random-dot stimuli used were designed to probe the robustness of V analysis by using specified outlier populations. Half of the dots in the right image were given a vertical magnification m , selected from the range 0–12%, and the other half were given a vertical magnification $-m/2$ in the range 0–6%. As m increased, the two populations of points were therefore consistent with increasingly different viewing geometries.

Figure 1 shows Monte-Carlo simulations of the performance of a non-robust and a highly robust ideal observer on the nulling task. For an ideal observer, the nulling magnification required can be shown to equal its estimate of the viewing parameter α , so estimated α values from both observers are plotted against magnification m

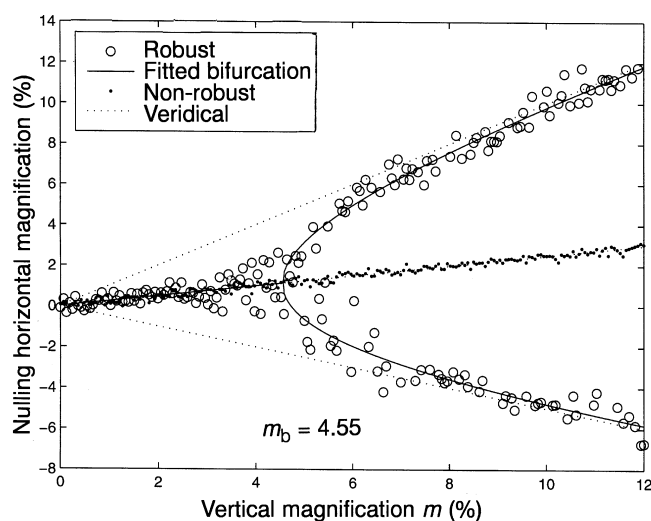


Figure 1 Monte-Carlo simulation of the performance of a non-robust and robust observer on the nulling task. The solid circles show the performance of the non-robust observer and the open circles show that of the highly robust ideal observer. The abscissa shows the range of m values used. The dotted lines show predicted induced effects m and $-m/2$ for single magnifications and for perfect pooling ($m/4$). The non-robust observer ($P = 1$, $k = 0.4^*$) pools over the whole range, whereas the robust observer ($P = 0.001$, $k = 0.4$) shows initial pooling behaviour followed by a bifurcation. Values of P and k_{30} were chosen to match the observed value of m_b .

for each trial. The non-robust observer sets the nulling horizontal magnifications to the mean magnification over the two populations for the whole range of m values; we describe this behaviour as pooling^{12,13}. In contrast, for the robust observer, initial pooling is followed by a bifurcation in which a magnification appropriate to one or other population is selected. Beyond a critical value of m , the robust observer in effect rejects one whole population as outliers and returns the viewing parameter appropriate to the other population. Which population is chosen is determined by statistical fluctuations.

Results for human observers are shown in Fig. 2 where the nulling magnifications recorded for each of four observers are plotted against m . As in the simulation, initial pooling is followed by a bifurcation at a critical magnification. A linear-hyperbolic bifurcation curve fitted the data significantly better than either a single straight line or two straight lines ($P < 0.001$ using a chi-squared statistic) and the fitted curve was used to estimate the critical magnification, m_b , which was found to be 4.4%. Consistent populations with magnifications of either m or $-m/2$ were also presented to the human observers. The group-mean settings for these stimuli are plotted with asterisks connected by lines. The two arms of the bifurcation for inconsistent data almost reach the curves for consistent populations. This indicates that human observers behave, at least qualitatively, like robust observers, and that data from one or other inconsistent population are successfully ignored for mixtures with large magnification differences. Note that the display appeared as a single planar entity throughout; observers did not report seeing two transparent planes at any time, as they do when populations with inconsistent H information are presented. This is consistent with the proposal that V is used to recover viewing parameters, which are intrinsically global in character. V is thus essentially different in character to H , which is a direct cue to shape and which is processed locally.

The measured critical value, m_b , offers the possibility of a quantitative comparison between human and robust ideal observer performance. To describe this comparison it is necessary to specify two aspects of the error model in more detail. First, acuity limitations were modelled as additive gaussian noise, the magnitude of which varied linearly with retinal eccentricity according to a model proposed in ref. 14. This model has a single free parameter, the

acuity at a chosen eccentricity. As the effects of vertical disparity are generally wide-field phenomena, we quote values for k_{30} , the retinal vertical acuity at 30° eccentricity. Second, the proportion of outliers is given by $1 - P$, where P is the probability that a given left-right feature match is correct. At a given retinal position these outlier disparity measurements are distributed uniformly over the feasible matching range.

We plotted m_b values for robust ideal observers with different combinations of k_{30} and P (Fig. 3). Although the dependence on k_{30} is strong and nearly linear there is also some dependence on P (in fact bifurcation can be delayed arbitrarily by taking P close enough to 1). In the region in which the observer could realistically be called robust, that is $P < 99\%$, the observed bifurcation at $m_b = 4.4\%$ is consistent with k_{30} values between 0.2° and 0.4°. Assuming that the human observer is robust and close to ideal, this suggests a value for k_{30} of around 0.3°. This is consistent with horizontal-disparity acuity for single points and with the value assumed for vertical acuity in ref. 15. Because no direct measurement of vertical-disparity acuity has been reported we regard our approach as providing an independent estimate derived from a task chosen for its suitability for solution using vertical-disparity information.

The robust ideal observer for small P is optimized to deal with situations in which most of the data are incorrect. This is the case when vertical-disparity analysis occurs before match verification using all potential matches within the allowed vertical-matching range. Contrary to what one might expect, performance in recovering viewing parameters does not show any pronounced decline in this situation. A robust ideal observer can reliably recover α and β and the equivalent g and d parameters on a data set containing a large majority of incorrect matches (Fig. 4). This is important because, if viewing parameters can be recovered before the matching problem is solved, they can greatly simplify its subsequent solution by restricting the search for correct matches to those consistent with the prevailing viewing geometry. We speculate that robust estimators may be able to perform other stereoscopic tasks without previously solving the matching problem. This may help to explain why the human visual system can sometimes tolerate considerable perturbations in the vertical locations of corresponding points in the left and right images^{16,17}.

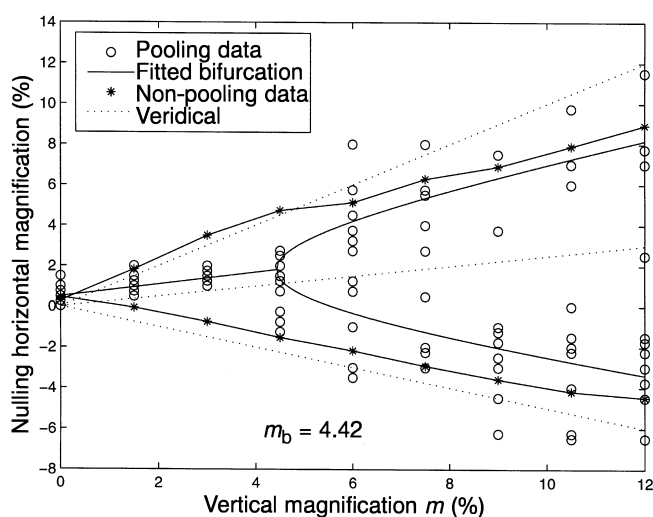


Figure 2 Pooling experiment where nulling magnifications for each of four observers are plotted against vertical magnification. The open circles show data for mixed magnification stimuli; they represent individual settings for each particular pairing of m and $-m/2$. The bifurcation from pooling behaviour is clearly visible. The best-fit linear-hyperbolic bifurcation curve ($m_b = 4.4\%$) is also shown. Asterisks connected by lines show group-mean data for consistent populations with vertical-image magnifications of either m (upper line) or $-m/2$ (lower line).

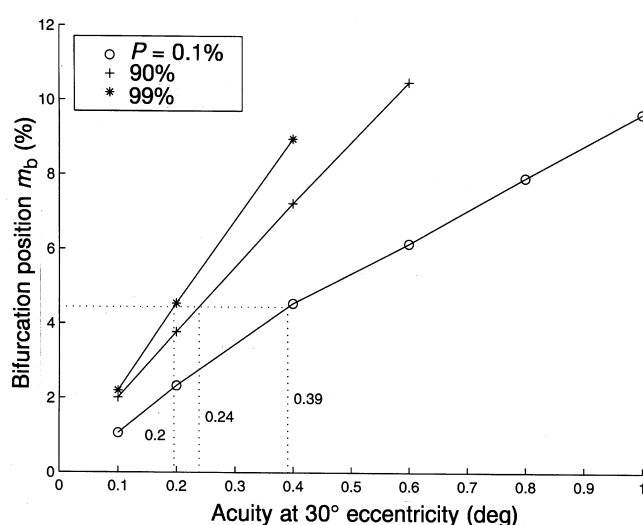


Figure 3 Bifurcation position versus acuity. The plotted curves show the predicted dependence of critical magnification on acuity k_{30} (abscissa) for three values of P , 0.1%, 90%, 99%, covering the range of 'robust' observers. The value of k_{30} giving the observed bifurcation position $m_b = 4.4\%$ is shown for each value of P .

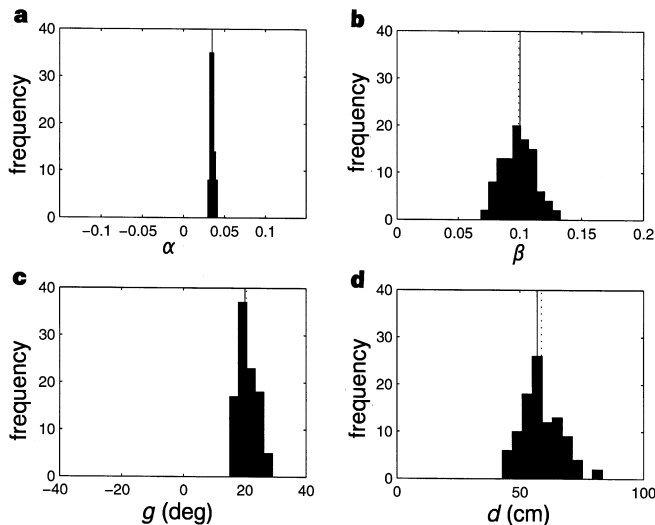


Figure 4 Performance of the robust ideal observer. The four graphs show histograms of 200 α and β values (**a** and **b** respectively) and the equivalent g and d values (**c** and **d** respectively) recovered by the robust ideal observer in the Hough limit ($P = 0.001$, $k = 0.4^\circ$) for a 20-cm diameter, frontoparallel surface patch at a distance of 57 cm. For every correct match (512 in all) there were 10 incorrect matches distributed uniformly over the possible vertical matching range at that location. Dotted lines show the true parameter value and solid lines show the population mean. The simulation shows usable accuracy and small bias for all parameters, with particularly good recovery of α .

The limit of small P has a further useful property. In the limit as $P \rightarrow 0$ the robust ideal observer can be implemented using an efficient parameter-estimation algorithm which is often used in computer vision and which is called the Hough transform^{18–20}. Hough algorithms are suitable for biological implementation as low-connectivity neural nets²¹. In our application, viewing-parameter pairs can be place-coded by a two-dimensional layer of neurons taking a weighted sum of inputs from disparity-selective binocular cells with which they are consistent. If these neuronal inputs measure the a-priori probability that a binocular feature has been correctly detected, the output will measure the posterior probability that the coded viewing parameters are correct. Winner-take-all competition in the output layer then estimates the optimal viewing parameters. Note that because this implementation codes probabilities, it is well adapted to the statistical combination of V information with viewing-parameter information from other sources (for example the oculomotor system).

We conclude by emphasizing three points. First, the psychophysical results show that human performance is well modelled by the robust ideal observer assumption, yielding a new method for estimating vertical-disparity acuity. Second, the analysis is based on the observation of a bifurcation onset rather than on a performance threshold. To our knowledge, an ideal observer analysis of human performance has not previously been based on a critical phenomenon of this kind. Third, we have derived a biologically plausible algorithm for vertical-disparity pooling which can recover accurate viewing-parameter estimates from noisy measurements of vertical disparity, including large numbers of incorrect matches. Thus the Bayesian approach, which would seem at first sight to offer information only about performance thresholds, can provide insights into perceptual systems at all three of the levels of understanding described in ref. 22: the computational (task description) level, the algorithmic (information representation and transformation) level and the implementation (physical realization or hardware) level. $\square$

Methods

Experiment. Stereograms were presented on a Silicon Graphics Indy monitor using Crystal Eyes stereo glasses. A chin-rest, bite-bar and forehead bar kept observers' heads square to the screen and 57 cm from it. The room was completely dark. The edges of the monitor and the keyboard were covered with black card to exclude external sources of V information. The stimuli were roughly circular patches (diameters subtended about 21.5 cm) of 512 red dots (dot size < 2 arcmin). Key-presses stepped horizontal magnification from -9 to 15% in steps of 0.25% , with a random starting magnification on each trial. Each observer produced three settings for each condition, collected from trials shown in randomized orders and distributed over three ~ 15 -min sessions separated by breaks of at least 4 h. Each condition was seen for ~ 60 s to ensure that the induced effect had reached a steady state²³. The four participants, one female and three males, were all experienced in psychophysical experiments. They were aged between 24 and 40 years and all had normal or corrected-to-normal vision and stereo acuity of 50 arcsec or better. Three participants (J.P., W.J.A. and D.B.) were authors; their data were not qualitatively different from those of the other observer.

Theory. The horizontal and vertical components of disparity of a point with retinal coordinates (x, y) and depth z relative to the fixation point in the local cyclopean frame are given to good accuracy by the linearized disparity equations of refs 7, 8:

$$H = \beta^2 \frac{z}{I} + \alpha x + \beta x^2$$

$$V = \alpha y + \beta xy$$

These equations can be extended to take account of the effects of eye elevation and cyclotorsion but we will not consider that refinement here¹⁰. Our hypothesis is that viewing parameters are recovered from V and used to recover scene structure z from H . Suppose that the true value of α is zero. Vertical magnification of the right image adds to V a term proportional to y , leading to a non-zero estimate for α . Allowing for the contribution of αx to H , the induced effect is generated. Horizontal magnification of the right image adds a term μx to H which exactly nulls this effect when $\mu = -\alpha$; hence the nulling task is a measurement of the viewing-system estimate of α .

If the correct viewing parameters are α and β , the vertical-disparity error on a match $M = (x, y, V)$ is $n = V - \alpha y - \beta xy$. The distribution of these errors is modelled as the sum of a uniform distribution due to incorrect matches plus a gaussian distribution representing acuity limitations:

$$p(n) = \frac{1-P}{\rho} + \frac{P}{\sigma\sqrt{2\pi}} \exp - \frac{n^2}{2\sigma^2} \propto 1 + q \exp - \frac{n^2}{2\sigma^2}$$

where $q = P/(1-P)(\rho/\sigma\sqrt{2\pi})$, ρ is the vertical matching range at that retinal position and σ is vertical-disparity acuity. Variation of σ with eccentricity E is modelled using the formula¹⁴

$$\sigma = \sqrt{2}k_0(1 + 0.59E)$$

where k_0 is foveal acuity, E is measured in degrees and the factor of $\sqrt{2}$ accounts for addition of independent errors in the two eyes.

Using Bayes' theorem the probability density function $p(\alpha, \beta | M_1, \dots, M_N)$ for the viewing parameters α and β , given the information in N matches, is

$$p(\alpha, \beta | M_1, \dots, M_N) \propto \prod_i p(m_i | \alpha, \beta) \propto \prod_i p(n_i) \propto \prod_i \left(1 + q_i \exp - \frac{n_i^2}{2\sigma_i^2} \right)$$

where we have assumed no previous knowledge and independence of match information. The ideal observer for α given the hit-or-miss²⁴ cost-function chooses α to maximize the marginal posterior density function

$$p(\alpha | M_1, \dots, M_N) = \int p(\alpha, \beta | M_1, \dots, M_N) d\beta$$

Alternatively, the ideal observer for both α and β maximizes $p(\alpha, \beta | M_1, \dots, M_N)$. The second procedure was used in the simulations. Maximization was performed numerically over a fine grid of α, β values and refined by interpolation. In the limit $P \rightarrow 0$, all the q_i values are small and we can approximate the product above by

$$p(\alpha, \beta | M_1, \dots, M_N) \propto 1 + \sum_i w_i q_i \quad \text{where} \quad w_i = \exp - \frac{n_i^2}{2\sigma_i^2}$$

Here each match makes an additive contribution q_i (which can be regarded as a measure of match goodness) weighted by its consistency w_i with a given α , β pairing. As the w_i values are small away from the line $\alpha x_i + \beta x_j y_i = V_i$ in α , β space, the algorithm reduces to a Hough transform^{18,20,25} where each match makes a line of contributions into an accumulator array.

Received 1 June; accepted 3 November 1998.

- Knill, D. C. & Richards, W. *Perception as Bayesian Inference* (Cambridge Univ. Press, Cambridge, 1996).
- Adelson, E. H. & Weiss, Y. A simple Bayesian model predicts a complex set of phenomena. *Perception* **27** (Suppl.), 18 (1980).
- Barlow, H. The absolute efficiency of perceptual decisions. *Phil. Trans. R. Soc. Lond. B* **290**, 71–82 (1980).
- Gårding, J., Porrill, J., Mayhew, J. E. W. & Frisby, J. P. Stereopsis, vertical disparity and relief transformations. *Vision Res.* **35**, 703–722 (1995).
- Koenderink, J. J. & van Doorn, A. J. Geometry of binocular vision and a model for stereopsis. *Biol. Cybern.* **21**, 29–35 (1976).
- Petrov, A. P. A geometrical explanation of the induced size effect. *Vision Res.* **20**, 409–416 (1980).
- Mayhew, J. E. W. The interpretation of stereo-disparity information: the computation of surface orientation and depth. *Perception* **11**, 387–403 (1982).
- Mayhew, J. E. W. & Longuet-Higgins, H. C. A computation model of binocular depth perception. *Nature* **297**, 376–379 (1982).
- Frisby, J. P. An old illusion and a new theory of stereoscopic depth perception. *Nature* **307**, 592–593 (1984).
- Porrill, J., Mayhew, J. E. W. & Frisby, J. P. in *Frontiers of Visual Science: Proceedings of the 1985 Symposium* 90–108 (National Academy Press, Washington DC, 1985).
- Ogle, K. N. in *The Eye* Vol. 4 Part II (ed. Davson, H.) 209–417 (Academic, New York, 1962).
- Stanton, P., Frisby, J. P. & Mayhew, J. E. W. Vertical disparity pooling and the induced effect. *Nature* **309**, 622–623 (1984).
- Adams, W. et al. Pooling of vertical disparity information in the human visual system. *Perception* **25**, 165–176 (1996).
- Drasdo, N. The neural representation of visual space. *Nature* **266**, 554–556 (1977).
- van de Grind, W. A., Erkelens, C. J. & Laan, A. C. Binocular correspondence and visual direction. *Perception* **24**, 215–235 (1995).
- Rogers, B. J. & Bradshaw, M. F. Does the visual system use the epipolar constraint for matching binocular images? *Invest. Ophthalmol. Vis. Sci.* **37**, 684 (1996).
- Stephenson, S. B. & Schor, C. M. Human stereo matching is not restricted to epipolar lines. *Vision Res.* **37**, 2717–2723 (1997).
- Stephens, R. S. The probabilistic Hough transform. *Image Vision Comput.* **9**, 66–71 (1991).
- Ballard, D. H. Generalising the Hough transform to detect arbitrary shapes. *Pattern Recog.* **13**, 111–122 (1981).
- Peck, S., Mayhew, J. E. W. & Frisby, J. P. Obtaining viewing distance and angle of gaze from vertical disparities using a Hough-type accumulator. *Image Vision Comput.* **2**, 180–190 (1984).
- Ballard, D. H. Cortical connections and parallel processing: structure and function. *Behav. Brain Sci.* **9**, 67–120 (1986).
- Marr, D. *Vision* (Freeman, San Francisco, 1982).
- Kaneko, H. & Howard, I. P. Spatial limitation of vertical-size disparity processing. *Vision Res.* **37**, 2871–2878 (1997).
- Kay, S. M. *Fundamentals of Signal Processing—Estimation Theory* (Prentice Hall International, New Jersey, 1993).
- Porrill, J. & Frisby, J. P. Computation of binocular eye position from vertical disparities with the use of probabilistic place coding. *Perception* **26**, 39 (1997).

Correspondence and requests for materials should be addressed to J.P.F.

The protein MAP-1B links GABA_C receptors to the cytoskeleton at retinal synapses

Jonathan G. Hanley, Peter Koulen*, Fiona Bedford, Phillip R. Gordon-Weeks† & Stephen J. Moss

MRC Laboratory for Molecular Cell Biology and Department of Pharmacology, University College, Gower Street, London WC1E 6BT, UK

* Max-Planck-Institut für Hirnforschung, Neuroanatomische Abteilung, Deutschordenstrasse 46, 60528 Frankfurt, Germany

† Developmental Biology Research Centre, Division of Biomedical Sciences, Kings College London, 26–29 Dury Lane, London WC2B 5RL, UK

The ionotropic type-A and type-C receptors for the neurotransmitter γ -aminobutyric acid (GABA_A and GABA_C receptors) are the principal sites of fast synaptic inhibition in the central nervous system^{1–3}, but it is not known how these receptors are localized at GABA-dependent synapses. GABA_C receptors, which are composed of ρ -subunits^{3–6}, are expressed almost exclusively in the retina of adult vertebrates, where they are enriched on

bipolar cell axon terminals^{7–9}. Here we show that the microtubule-associated protein 1B (MAP-1B) specifically interacts with the GABA_C $\rho 1$ subunit but not with GABA_A receptor subunits. Furthermore, GABA_C receptors and MAP-1B co-localize at post-synaptic sites on bipolar cell axon terminals. Co-expression of MAP-1B and the $\rho 1$ subunit in COS cells results in a dramatic redistribution of the $\rho 1$ subunit. Our observations suggest a novel mechanism for localizing ionotropic GABA receptors to synaptic sites. This mechanism, which is specific for GABA_C but not GABA_A receptors, may allow these receptor subtypes, which have distinct physiological and pharmacological properties, to be differentially localized at inhibitory synapses.

To investigate the mechanisms responsible for the localization of ionotropic GABA receptors in neurons, we used a yeast two-hybrid screen¹⁰. We focused on GABA_C receptors, which are composed of $\rho(1–3)$ subunits, unlike the more structurally diverse GABA_A receptors, which are assembled from five subunit classes: $\alpha(1–6)$, $\beta(1–4)$, $\gamma(1–4)$, δ and ϵ (refs 1, 2). A retinal complementary DNA library was screened with a bait encoding the large intracellular domain of the GABA_C-receptor $\rho 1$ subunit^{1–3}. One of the strongly interacting clones, clone 8 (Fig. 1), corresponds to amino acids 460–690 of MAP-1B, which is 95% identical to this portion of the human protein (Fig. 1b)^{11–14}. This interaction must be specific to the $\rho 1$ subunit as clone 8 does not interact with the intracellular domains of selected GABA_A-receptor subunits^{1,2}, nor with that of the GABA_C-receptor $\rho 2$ subunit⁶ (Fig. 1a). These constructs all produced fusion proteins with Gal-4 in yeast of the expected sizes (relative molecular masses (M_r) were determined by western blotting; data not shown). To pinpoint the binding site for the $\rho 1$ subunit on MAP-1B, smaller sections of clone 8 were tested for their interaction in yeast. Binding of the $\rho 1$ subunit was critically dependent on amino-acid residues 460–565, but independent of the putative microtubule-binding domain within clone 8 (Fig. 1c)^{11–14}. The interaction between clone 8 and the intracellular domain of the $\rho 1$ subunit was confirmed by expressing both proteins as glutathione-S-transferase (GST) fusions¹⁵. Gel-overlay assay indicated that GST- $\rho 1$ binds to GST-clone 8 but not to GST alone (Fig. 1d).

We analysed the interaction between the $\rho 1$ subunit and MAP-1B

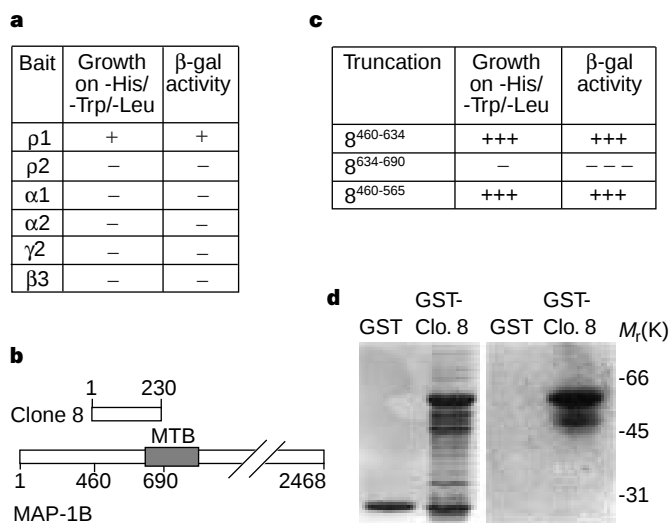


Figure 1 Specific interaction of the $\rho 1$ intracellular domain with an N-terminal domain of MAP-1B. **a**, Interaction of clone 8 and the intracellular domain of the $\rho 1$ subunit in yeast, as measured by growth on -His/-Trp/-Leu medium and β -galactosidase activity. **b**, Clone 8 encodes amino acids 460–690 of MAP-1B. **c**, The domain in MAP-1B responsible for binding $\rho 1$, as determined by deletion analysis in yeast; interaction was measured as for **a**. **d**, ³²P-labelled GST- $\rho 1$ interacts with GST-clone 8 (clo.8) in an overlay assay. Right, autoradiography; left, identical gel stained with Coomassie brilliant blue.

in rat retina by using specific antibodies. Anti- $\rho 1$ serum recognized a protein of M_r 45K in both retinal extracts and in COS cells transiently expressing the $\rho 1$ subunit (Fig. 2a). Anti-MAP-1B recognized a major band of 330K in retina¹⁶ and in COS cells transiently expressing MAP-1B (Fig. 2b). Retinal extracts were immunoprecipitated using anti-MAP-1B and then probed with anti- $\rho 1$. The $\rho 1$ subunit was precipitated with anti-MAP-1B, but not with control sera (Fig. 2c). Preabsorption of anti-MAP-1B with antigen¹⁶ also prevented precipitation of the $\rho 1$ subunit (Fig. 2c). In contrast, the GABA_A-receptor $\gamma 2$ subunit, which is an essential component of benzodiazepine-sensitive GABA_A receptors¹⁷, could not be precipitated with MAP-1B (Fig. 2c). Interaction was also analysed by affinity purification using 'pull-down' assays, in which MAP-1B bound to GST- $\rho 1$ but not to GST-fusion proteins containing the intracellular domains of a range of GABA_A-receptor subunits, or to GST alone (Fig. 2d). This interaction was specific to MAP-1B as the homologous microtubule-associated protein MAP-1A (M_r 350K) did not bind to GST- $\rho 1$ (Fig. 2d). The binding site on $\rho 1$ for MAP-1B was likewise analysed. Using a range of GST- $\rho 1$ constructs, amino acids 434–454 were found to be critical for binding MAP-1B (Fig. 2e). As amino acids 446–454 are identical in both $\rho 1$ and $\rho 2$, but $\rho 2$ does not bind MAP-1B (Fig. 1), these results indicate that residues 435–446 in $\rho 1$ are probably sufficient to bind MAP-1B.

As MAP-1B will bind to both actin and tubulin^{11–14,18}, we analysed whether MAP-1B/ $\rho 1$ complexes could interact with either of these components of the cytoskeleton by using pull-down assays. Both actin and tubulin bound to GST- $\rho 1$ in retinal extracts with MAP-1B, but not in extracts prepared from COS cells (Fig. 2f). As COS cells do not express significant amounts of MAP-1B, this suggests that the $\rho 1$ subunit does not interact directly with either actin or tubulin (Fig. 2h, f). Together, these results indicate that MAP-1B links the GABA_A receptor to the cytoskeleton.

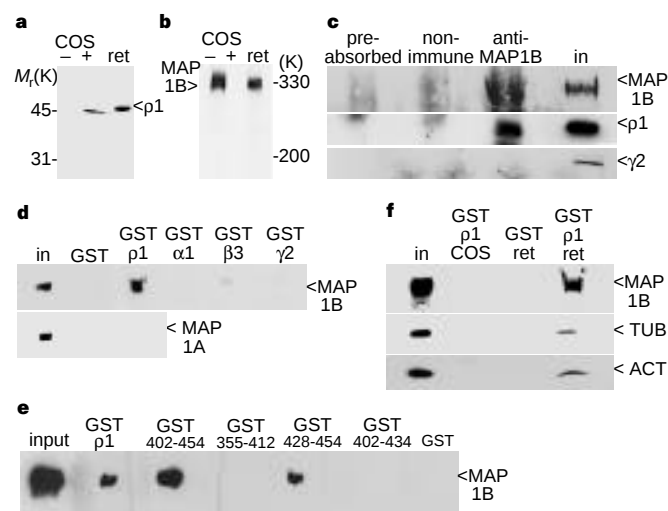


Figure 2 Interaction of MAP-1B with the GABA_A receptor $\rho 1$ subunit in retinal extracts. **a, b**, Characterization of anti- $\rho 1$ and anti-MAP-1B antisera by western blotting. **c**, Co-immunoprecipitation of the $\rho 1$ subunit, but not the $\gamma 2$ subunit, with MAP-1B; 'in', 1% of input. **d–f**, Pull-down assays using retinal extracts. **d**, GST- $\rho 1$, but not GST- $\alpha 1$, - $\beta 3$ or - $\gamma 2$ interacts with MAP-1B (top); MAP-1B but not MAP-1A binds to $\rho 1$ (bottom); 'in', 2% of input. **e**, Identification of the MAP-1B binding site on GST- $\rho 1$ using deletion analysis. **f**, Tubulin and actin bind to $\rho 1$ in extracts prepared from retina (ret) but not COS cells; 'in', 2% for MAP-1B, 0.1% for actin and tubulin of input. In **c–f**, co-precipitating proteins or binding partners were detected by western blotting.

We next compared the distribution of GABA_A receptors and MAP-1B in sections of adult rat retinae (Fig. 3a). Co-localization was evident in many layers of the retina, particularly in the synapse-rich inner plexiform layer, a major site of GABA_A-receptor expression^{8,9}. Co-localization was less marked in the outer retina, which may reflect the presence of GABA_A receptors that do not contain the $\rho 1$ subunit¹⁹, or alternatively MAP-1B may perform a different function in this region of the retina^{11,12}. Multiple functions have also been inferred for the NMDA-receptor-anchoring protein α -actinin-2 and the glycine-receptor-clustering protein gephyrin^{20,21}. We analysed the localization at the single-cell level in dissociated bipolar neurons: co-localization of MAP-1B and ρ -subunit immunoreactivity was evident in bipolar cell axon terminals (Fig. 3b), which form axo-axonic synapses with GABAergic amacrine neurons²².

To investigate how MAP-1B controls the subcellular localization of GABA_A receptors, we expressed the $\rho 1$ subunit and MAP-1B in COS cells (Fig. 4). Expression of the $\rho 1$ subunit by itself produced diffuse surface and intracellular staining, as measured by using anti- ρ serum (Fig. 4a). MAP-1B expression was similarly analysed using the monoclonal antiserum AA6, and under our fixation conditions showed a diffuse intracellular staining¹⁴ (Fig. 4b). Co-expression of MAP-1B with the $\rho 1$ subunit resulted in a dramatic change in the distribution of $\rho 1$ into a more punctate intracellular localization (Fig. 4c, d). Co-expression of gephyrin and glycine receptors in heterologous systems also leads to the re-direction of these receptors into intracellular aggregates²³, whereas in neurons, gephyrin mediates surface-receptor clustering²⁴. As a control, we examined the effects of MAP-1B expression on functional GABA_A receptors composed of $\beta 3$ subunits²⁵. We found that MAP-1B did not effect the distribution of the $\beta 3$ subunit in COS cells (Fig. 4e–g). The ability of MAP-1B to trigger the aggregation of $\rho 1$ subunits indicates that this protein may be important for controlling the localization and clustering of GABA_A receptors in neurons.

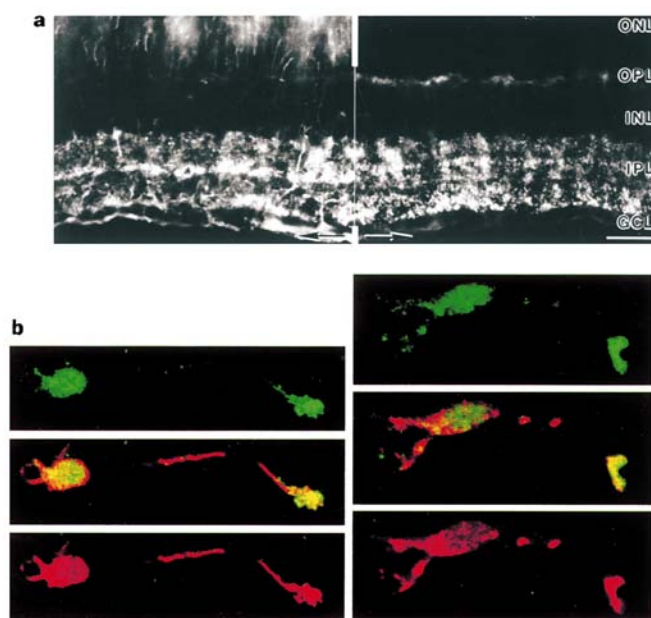
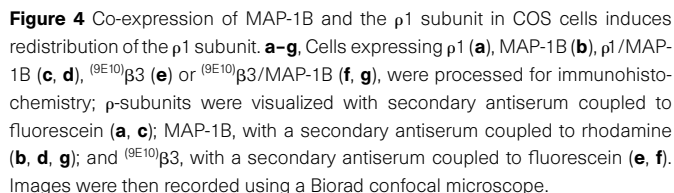


Figure 3 Co-localization of ρ -subunits and MAP-1B immunoreactivity in the retina. **a**, Retinal sections, or **b**, dissociated retinal bipolar cells, were fixed and processed for immunohistochemistry. In **a**, ρ -subunits were visualized with Cy3 (right), and MAP-1B with fluorescein (left). In the mirror images, symmetry across the midline represents co-localization. ONL, outer nuclear layer; OPL, outer plexiform layer; INL, inner nuclear layer; IPL, inner plexiform layer; GCL, ganglion cell layer; scale bar, 25 μ m. In **b**, ρ -subunits were visualized with fluorescein, and MAP-1B with rhodamine. Images were recorded from two bipolar cells with distinct morphologies; merged images are shown in the central panels. Scale bar, 5 μ m.



Our observations suggest a mechanism by which ionotropic GABA receptors are linked to the cytoskeleton at synaptic sites. This mechanism, which is mediated by MAP-1B, is specific for GABA_C but not GABA_A receptors, and may underlie the differing subcellular distribution of these distinct ionotropic GABA receptors in retinal neurons⁹. Given that these receptor subtypes have differing physiological and pharmacological properties^{2,3}, this mechanism may be important in facilitating fast inhibitory neurotransmission in the central nervous system. □

Yeast two-hybrid screens. Approximately 3 million clones were screened from a bovine retinal cDNA library constructed in the Gal4 activation-domain-carrying vector pACTII (Clontech). The library was screened with a bait encoding the intracellular domain of the human $\rho 1$ subunit⁷ cloned in the Gal4-DNA-binding-domain vector pAS2-1 (Clontech). The plasmids were transformed into yeast strain Y190 and transformants were selected on triple dropout media (-Leu/His/Trp containing 30 mM 3-amino-triazole) and assayed for β -galactosidase activity¹⁰. Positive clones were co-transformed with either the bait vector or the pACTII vector backbone into yeast in order to confirm interactions. Baits encoding the GABA_A receptor $\alpha 1$, $\alpha 2$, $\beta 3$, $\gamma 2$ (refs 1, 2) and $\rho 2$ (ref. 3) subunits in pAS2-1 were screened against clone 8 in pACT II (ref. 10). All of these constructs produced Gal4 fusion proteins of the predicted sizes on expression in yeast, as measured by western blotting using anti-Gal4 antiserum (Clontech). To map the region of MAP-1B responsible for binding the intracellular domain of the $\rho 1$ subunit, fragments of clone 8 were amplified by PCR and cloned into pACTII. Interactions with the intracellular domain of the $\rho 1$ subunit were then measured in yeast as described. All constructs used were produced by cloning PCR products in-frame into the appropriate vectors and verification by sequencing.

(anti-p) was used that recognizes the p1–3 subunits^{8,9}. Monoclonal antibody AA6 (Sigma) against MAP-1B (ref. 14) was used for histochemical analysis. A rabbit polyclonal antiserum against MAP-1B (anti-MAP-1B)¹⁶ was used for immunoprecipitation and blotting. The GABA_A-receptor $\gamma 2$ subunit was detected using an affinity-purified antiserum raised against a synthetic peptide corresponding to the first 29 residues of this protein.

Immunohistochemistry. Retinae were removed from halothane-anaesthetized adult (P64) rats and processed for fluorescence as described^{8,9}. MAP-1B was detected using a monoclonal antibody against rat MAP-1B (ref. 10) (anti-MAP-1B AA6; sigma) and visualized using a fluorescein-conjugated secondary antiserum. ρ -Subunits were detected using rabbit anti- ρ sera^{8,9} and visualized using a secondary antiserum coupled to Cy3 (Dianova, Germany). Images were collected from each channel and then printed as mirror images, cut and aligned along a common border. The black arrows around the midline in Fig. 3a indicate the orientation of the mirror images. The specificity of the double-labelling experiments was confirmed by omitting primary antisera. Images were collected on a Zeiss photomicroscope. Dissociated retinal cells were prepared from adult rats and processed for immunohistochemistry as described⁸; retinal cultures were then examined using confocal microscopy. Fluorescence signals could only be seen in bipolar neurons, which account for 1–2% of the total cell population, confirming the specificity of the antisera.

Cell transfection and immunofluorescence. COS cells were transfected with expression constructs by electroporation^{25,26}. Expression of the $\rho 1$ subunit and the GABA_A-receptor $\beta 3$ subunit (modified with the 9E10 epitope²³) was under the control of the cytomegalovirus promoter. MAP-1B expression was driven by the SV40 promoter. 48 h after transfection, cells were fixed using 3% paraformaldehyde, permeabilized with 0.2% N-P40 and processed for immunofluorescence^{25,26}. The $\rho 1$ subunit was detected with rabbit anti- ρ , MAP-1B was detected with monoclonal antiserum AA6, and the GABA_A-receptor (9E10) $\beta 3$ -subunit was detected using rabbit anti-9E10 sera. The specificity of the staining was established by omitting the primary antiserum in each case. Also, no staining was seen with either antiserum in untransfected cells. Images were recorded by confocal microscopy.

1. Macdonald, R. L. & Olsen, R. W. GABA_A receptor channels. *Annu. Rev. Neurosci.* **17**, 569–602 (1994).
2. Rabow, L. E., Russek, S. J. & Farb, D. H. From ion currents to genomic analysis: recent advances in GABA_A receptor research. *Synapse* **21**, 189–274 (1995).
3. Lukaszewicz, P. D. GABA_C receptors in the vertebrate retina. *Mol. Neurobiol.* **12**, 181–194 (1996).
4. Polenzani, L., Woodward, R. M. & Milei, R. Expression of mammalian γ -aminobutyric acid receptors with distinct pharmacology in *Xenopus*. *Proc. Natl. Acad. Sci. USA* **88**, 4318–4322 (1991).
5. Cutting, G. R. *et al.* Cloning of the γ -aminobutyric acid (GABA) rho1 cDNA: aGABA receptor subunit highly expressed in the retina. *Proc. Natl. Acad. Sci. USA* **88**, 2673–2677 (1991).
6. Wang, T. L., Guggino, W. B. & Cutting, G. R. A novel γ -aminobutyric acid receptor subunit (rho2) cloned from human retina forms bicuculline-insensitive homooligomeric receptors in *Xenopus* oocytes. *J. Neurosci.* **14**, 6524–6531 (1994).
7. Enz, R., Brandstatter, J. H., Hartveit, E., Wassle, H. & Bormann, J. Expression of GABA receptor rho1 and rho2 subunits in the retina and brain of the rat. *Eur. J. Neurosci.* **7**, 1495–1501 (1995).
8. Enz, R., Brandstatter, J. H., Hartveit, E., Wassle, H. & Bormann, J. Expression of GABA receptor rho1 and rho2 subunits in the retina and brain of the rat. *Eur. J. Neurosci.* **7**, 1495–1501 (1995).
9. Koulen, P., Brandstatter, J. H., Enz, R. & Wassle, H. Synaptic clustering of GABA_C receptor ρ -subunits in the rat retina. *Eur. J. Neurosci.* **10**, 115–127 (1998).
10. Field, S. & Song, O. A novel genetic system to detect protein–protein interactions. *Nature* **340**, 245–246 (1989).
11. Hammarback, J. A. in *Brain Microtubule-associated Proteins: Modifications in Disease* (eds Avila, J., Kosik, K. & Brandt, R.) 1–17 (Harwood, Amsterdam, 1997).
12. Ulloa, L., Avila, A. & Diaz-Nido, R. in *Brain Microtubule-associated Proteins: Modifications in Disease* (eds Avila, J., Kosik, K. & Brandt, R.) 17–33 (Harwood, Amsterdam, 1997).
13. Lien, L., Feener, C. A., Fischbach, N. & Kunkel, L. M. Cloning of human microtubule-associated protein 1B and the identification of a related gene on chromosome 15. *Genomics* **22**, 273–280 (1994).
14. Noble, M., Lewis, S. A. & Cowan, N. The microtubule binding domain of microtubule-associated protein MAP-1B contains a repeated sequence motif unrelated to that of MAP2 and tau. *J. Cell. Biol.*

- 109, 3367–3376 (1989).
15. Smith, D. B. & Johnson, K. S. Single-step purification of polypeptides expressed in *E. coli* as fusions with glutathione S-transferase. *Gene* **67**, 31–40 (1988).
 16. Johnstone, M., Goold, R. G., Fischer, I. & Gordon-Weeks, P. R. The neurofilament antibody RT97 recognises a developmentally regulated phosphorylation epitope on microtubule-associated protein 1B. *J. Anat.* **191**, 229–244 (1997).
 17. Gunther, U. *et al.* Benzodiazepine-insensitive mice generated by targeted disruption of the $\gamma 2$ subunit of GABA_A receptors. *Proc. Natl Acad. Sci. USA* **92**, 7749–7753 (1995).
 18. Fujii, T., Watanabe, M., Ogoma, Y., Kondo, Y. & Arai, T. Microtubule-associated proteins, MAP 1A and MAP 1B, interact with F-actin *in vitro*. *J. Biochem.* **6**, 827–829 (1993).
 19. Albrecht, B. E. & Darlison, M. G. Localisation of the $\rho 1$ and $\rho 2$ subunit messenger RNAs in chick retina by *in situ* hybridization predicts the existence of γ -aminobutyric acid type C receptor subtypes. *Neurosci. Lett.* **189**, 155–158 (1995).
 20. Wyszynski, M. *et al.* Differential regional expression and ultrastructural localization of α -actinin-2, a putative NMDA-receptor-anchoring protein, in rat brain. *J. Neurosci.* **18**, 1383–1392 (1998).
 21. Prior, P. *et al.* Primary structure and alternative splice variants of gephyrin, a putative glycine receptor-tubulin linker protein. *Neuron* **8**, 1161–1170 (1992).
 22. Tachibana, M. & Kaneko, T. γ -Aminobutyric acid exerts a local inhibitory action on the axon terminal of bipolar cells: evidence for negative feedback from amacrine cells. *Proc. Natl Acad. Sci. USA* **85**, 3501–3505 (1987).
 23. Kirsch, J., Kuhse, J. & Betz, H. Targeting of glycine receptor subunits to gephyrin-rich domains in transfected human embryonic kidney cells. *Mol. Cell. Neurosci.* **6**, 450–461 (1995).
 24. Kirsch, J., Wolters, I., Triller, A. & Betz, H. Gephyrin antisense oligonucleotides prevent glycine-receptor clustering in spinal neurons. *Nature* **366**, 745–748 (1993).
 25. Connolly, C. N., Wooltorton, J. A., Smart, T. G. & Moss, S. J. Subcellular localization of GABA_A receptors is determined by receptor β subunits. *Proc. Natl Acad. Sci. USA* **93**, 9899–9904 (1996).
 26. Connolly, C. N., Krishek, B., McDonald, B. M., Smart, T. G. & Moss, S. J. Assembly and cell surface expression of heteromeric and homomeric γ -aminobutyric acid type A receptors. *J. Biol. Chem.* **271**, 89–96 (1996).
 27. Li, M., Jan, Y. & Jan, L. Specification of subunit assembly by the hydrophilic amino-terminal domain of the Shaker potassium channel. *Science* **257**, 1225–1229 (1992).

Acknowledgements. We thank C.-H. Sung for the retinal cDNA library, D. Attwell for help with retinal cultures, and W. Wisden and T. Smart for their comments on the manuscript. This work was supported by the MRC, the Wellcome Trust and the Royal Society.

Correspondence and requests for materials should be addressed to S.J.M. (e-mail: steve.moss@UCL.ac.uk).

GABA_A-receptor-associated protein links GABA_A receptors and the cytoskeleton

Hongbing Wang*, Fiona K. Bedford†, Nicholas J. Brandon†, Stephen J. Moss† & Richard W. Olsen*

*Molecular Biology Institute, University of California, Los Angeles, California, 90095 USA

†Medical Research Council Laboratory of Molecular and Cell Biology and the Department of Pharmacology, University College, London WC1E 6BT, UK

‡Department of Molecular & Medical Pharmacology, UCLA School of Medicine, Los Angeles, California 90095-1735, USA

Type-A receptors for the neurotransmitter GABA (γ -aminobutyric acid) are ligand-gated chloride channels that mediate inhibitory neurotransmission. Each subunit of the pentameric receptor protein has ligand-binding sites in the amino-terminal extracellular domain and four membrane-spanning regions, one of which forms a wall of the ion channel¹. Each subunit also has a large intracellular loop that may be a target for protein kinases and be required for subcellular targeting and membrane clustering of the receptor, perhaps by anchoring the receptor to the cytoskeleton^{2–4}. Neurotransmitter receptors need to be positioned in high density in the cell membrane at sites postsynaptic to nerve terminals releasing that neurotransmitter. Other members of the superfamily of ligand-gated ion-channel receptors associate in post-synaptic-membrane clusters by binding to the proteins rapsyn or gephyrin^{5–7}. Here we identify a new cellular protein, GABA_A-receptor-associated protein (GABARAP), which can interact with the $\gamma 2$ subunit of GABA_A receptors. GABARAP binds to GABA_A receptors both *in vitro* and *in vivo*, and co-localizes with the punctate staining of GABA_A receptors on cultured cortical neurons. Sequence analysis shows similarity between GABARAP and light chain-3 of microtubule-associated proteins 1A and 1B. Moreover, the N terminus of GABARAP is highly positively

charged and features a putative tubulin-binding motif. The interactions among GABA_A receptors, GABARAP and tubulin suggest a mechanism for the targeting and clustering of GABA_A receptors.

In an attempt to identify molecules that can mediate the association between GABA_A receptors and cellular proteins, including the cytoskeleton, we used the yeast two-hybrid system to screen for such gene products⁸. We obtained several clones from a human fetal brain complementary DNA library by using the intracellular loop of the

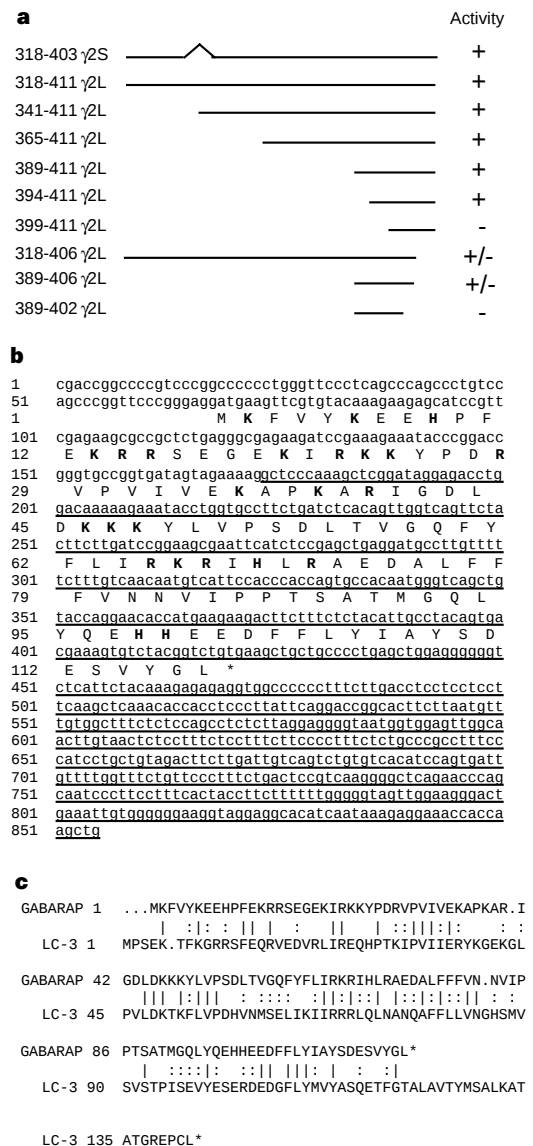


Figure 1 Segment of the GABA_A-receptor $\gamma 2$ subunit intracellular loop that is required for interaction with GABARAP, and sequence of GABARAP. **a**, The intracellular loops of $\gamma 2L$ and $\gamma 2S$ and different truncations were tested in the yeast two-hybrid assay for interaction with GABARAP. Amino-acid numbers are indicated on the left. +, activation of both *LacZ* and *LEU2*; +/-, activation of *LEU2* only; -, no gene activation. One LexA-binding site is present upstream of *LacZ*; four LexA-binding sites are present upstream of *LEU2*. **b**, The 855-base-pair cDNA of human GABARAP (lower-case letters) encodes a 117-amino-acid polypeptide (upper-case letters). Basic amino acids are shown in bold. The asterisk indicates the point at which translation stops. The original cDNA isolated in yeast two-hybrid screen is underlined. **c**, Sequence comparison of GABARAP and light chain-3 (LC-3) of MAP-1A and 1B. The amino-acid sequences of GABARAP and LC-3 were compared with the BestFit program of the GCG package. Identical amino acids are indicated by a vertical bar, and similar ones by two vertical dots. Dots in the sequences represent gaps introduced to maximize alignment, and asterisks represent the stop points of translation.

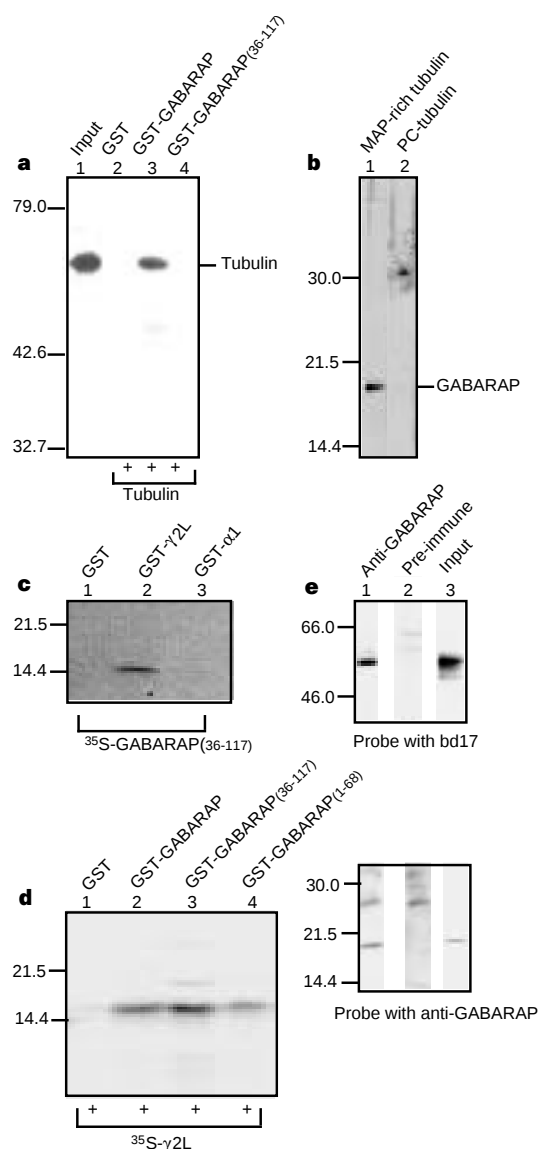


Figure 2 Interaction between GABA_A receptors, GABARAP and tubulin. **a**, Immunostaining of tubulin recovered from glutathione-agarose charged with GST (lane 2), GST-GABARAP (lane 3) and GST-GABARAP(36-117) (lane 4). Tubulin (1 µg) was loaded in lane 1. **b**, Co-purification of GABARAP with tubulin. MAP-rich tubulin (from Cytoskeleton; lane 1) was purified from cow brain by three cycles of heat-induced polymerization and cold-induced depolymerization. Further purification from MAP-rich tubulin with phosphocellulose (PC) to remove MAPs results in PC-tubulin. GABARAP was probed by anti-GABARAP antibody, followed by SDS-PAGE. **c**, Autoradiograph of ³⁵S-labelled GABARAP(36-117) recovered from glutathione-agarose charged with GST (lane 1), GST-γ2L (lane 2) or GST-α1 (lane 3). **d**, Autoradiograph of ³⁵S-labelled γ2L recovered from glutathione-agarose charged with GST (lane 1), GST-GABARAP (lane 2), GST-GABARAP(36-117) (lane 3) or GST-GABARAP(1-68) (lane 4). **e**, Co-immunoprecipitation of GABA_A receptors with GABARAP. Proteins precipitated with anti-GST-GABARAP antibody (lane 1) or pre-immune serum (lane 2) were analysed by western blot using bd17 (top) or antibodies against GST-GABARAP (bottom). Solubilized membrane used for co-immunoprecipitation was loaded in lane 3 as input. Co-immunoprecipitation using anti-GABARAP gave the same results. Anti-GABARAP absorbed by an antigen-affinity column did not precipitate GABARAP and GABA_A receptors (data not shown). No signal was detected in western blot analysis when primary antibodies were eliminated (data not shown). Numbers to the left of gels indicate relative molecular mass (in thousands).

GABA_A-receptor γ2S subunit or α1 subunit as bait. One clone (GABARAP) that interacted specifically with the γ2 subunit was further characterized. The γ2 subunit is not only the most abundant GABA_A-receptor subunit in the central nervous system⁹, but is also a substrate for several protein kinases^{2,10}. Activation of reporter genes in the yeast two-hybrid system was observed only when the isolated GABARAP cDNA was co-expressed with the γ2S- or γ2L-subunit intracellular loop, and was not observed when the cDNA was expressed with an unrelated *Drosophila Bicoid* product, with the β1 subunit or with several α-subunits; analysis using truncated forms of the γ2-subunit intracellular loop indicated that the carboxy-terminal 18 amino acids (R394 to D411) was both sufficient and necessary for the interaction (Fig. 1a). The amino-acid sequence within this portion is highly conserved among γ-subunits (83.3% identity and 88.9% similarity between γ2 and γ1 subunits; 50% identity and 83.3% similarity between γ2 and γ3 subunits) but much more diverse in other subtypes, indicating that GABARAP may be specific to γ-subunits of the receptor.

The full-length cDNA of GABARAP was obtained by screening a human adult brain λgt11 cDNA library. It encodes a 117-amino-acid polypeptide with a predicted relative molecular mass of 13,900 (*M_r* 13.9K) (Fig. 1b). The original isolate encodes amino acids 36 to 117 of GABARAP. A single messenger RNA transcript of ~900 bases was detected by northern blot analysis of human tissues, indicating that the 855-base-pair GABARAP cDNA is a full-length clone (data not shown).

We found no exact match of GABARAP in the GenBank database, so it must represent a new protein. However, it has sequence similarity to light chain-3 of microtubule-associated proteins (MAPs) 1A and 1B (31% identity and 64% similarity)¹¹ (Fig. 1c). Light chain-3 co-polymerizes with tubulin and is a component of the MAP-1 complex, which is composed of light chains 1, 2 and 3 and heavy chains. The sequence similarity between light chain-3 and GABARAP indicates that GABARAP may be another MAP or a component of certain MAP complexes. Like MAPs, GABARAP has an asymmetric structure, with a basic N-terminal region and an acidic C terminus, and an overall positive charge (an isoelectric point of 9.6). The highly positively charged regions in MAPs may mediate their binding to the acidic C-terminal tails of tubulin through ionic interactions, possibly by helix-helix interactions^{12,13}. Although no tubulin-binding motif identical to that of other MAPs is present in GABARAP, the N-terminal 21 amino acids could fold into an α-helix, according to Garnier-Osguthorpe-Robson prediction¹⁴. Interestingly, five basic amino acids (K2, K6, H9, K13 and K20) of GABARAP align on one side of the helix, with four out of the five amino acids appearing consecutively in this alignment. The α-helix formed by the N-terminal 21 amino acids may confer most of the tubulin-binding activity in GABARAP.

To determine whether GABARAP and tubulin interact, we used *in vitro* binding assays. Tubulin interacted with a fusion protein containing glutathione-S-transferase (GST) and GABARAP, but not with GST-GABARAP(36-117) or with GST (Fig. 2a). These results indicate that the N-terminal 36-amino-acid sequence may be necessary for tubulin-binding activity. Furthermore, like other MAPs, GABARAP was detected in a tubulin preparation (MAP-rich tubulin) purified by several cycles of polymerization and depolymerization (Fig. 2b). The presence of GABARAP in MAP-rich tubulin indicates its *in vivo* association with microtubules. Surprisingly, when we co-expressed full-length GABARAP and the γ2 GABA_A-receptor subunit in a yeast two-hybrid assay, we did not observe activation of reporter genes (data not shown). The lack of activation may result from the presence of the N-terminal tubulin-binding domain in the full-length GABARAP. Binding of GABARAP to microtubules in the cytoplasm may prevent it from entering the nucleus, and therefore activation of the reporter genes *LacZ* and *LEU2* will not occur.

In agreement with the yeast two-hybrid tests, ³⁵S-labelled

GABARAP(36–117) interacted with GST- γ 2L, but not with GST- α 1 or with GST (Fig. 2c). GABARAP(36–117) also interacted with γ 2S (data not shown). We also tested the association of γ 2L with different versions of GABARAP (Fig. 2d). Although not detected in the yeast system, full-length GABARAP did associate with the γ 2 subunit in the *in vitro* binding assay. The γ 2 subunit also interacted with GABARAP(1–68), indicating that the sequence between amino acids 36 and 68 of GABARAP is important for its GABA_A-receptor-binding activity.

We studied the association of native GABA_A receptors with GABARAP *in vivo* by co-immunoprecipitation, using solubilized membranes from rat brain. GABA_A receptors (β 2/3 subunit, M_r 56K) and GABARAP (M_r 17K) were both present in the precipitate brought down by antibodies against GABARAP (Fig. 2e). To investigate whether GABARAP co-localizes with GABA_A receptors *in vivo*, we double-stained primary cultures of rat cerebral cortex with antibodies against GABARAP and GABA_A receptors. Punctate staining of GABA_A receptors was found in both cell bodies and processes of numerous neurons (Fig. 3b), and a similar pattern was observed for GABARAP (Fig. 3a). The clusters formed by GABARAP co-localized with those formed by GABA_A receptors (Fig. 3c). This punctate staining was observed in 10–20% of the cells in the culture, although some cells were not stained and others stained diffusely. Antibodies against the GABA_A-receptor α 1 and β 2/3 subunits have been reported to stain principle neurons in rat cortex hippocampus diffusely and sharply outlined cells with puncta were limited to a subpopulation of GABA-responsive interneurons^{4,15}. The co-localization of GABARAP with GABA_A receptors in clusters in cultured neurons, presumably at synapses^{4,15}, indicates functional interaction of GABARAP with GABA_A receptors and a likely role of GABARAP in clustering.

We determined the expression pattern of GABARAP by northern blot and western blot analysis. GABARAP mRNA was detected in all tissues tested, namely heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas (data not shown). We measured the amount of protein expressed by using antibodies against GST-GABARAP. A molecule of M_r 17K was detected; this molecule co-migrated with the recombinant GABARAP, providing further evidence that the GABARAP cDNA that we isolated is a full-length clone. In agreement with the northern blots, GABARAP protein was expressed in all tissues tested as well as in different brain subregions (data not shown). The wide expression suggests the involvement of GABARAP in biological events other than interaction with GABA_A receptors. This wide distribution of expression is also observed for the glycine-receptor linker protein gephyrin⁷.

Rapsyn and gephyrin may be involved in the clustering, stability and mobility of receptors of the superfamily of ligand-gated ion-channel receptors^{5,7,16}. Phosphorylation of the linker protein-

receptor complex may be important in synaptic plasticity mechanisms^{17,18}. Actin, tubulin and other associated proteins co-purify with GABA_A receptors^{19,20}, but no linker proteins had, to our knowledge, been identified until now. The identification of GABARAP opens an avenue for studies of the regulation, targeting, clustering, degradation and internalization of GABA_A receptors, as well as of the significance of phosphorylation on synaptic plasticity and interactions with the cytoskeleton. □

Methods

Yeast two-hybrid screening. To generate the bait plasmids, pEG202- γ 2S, pEG202- γ 2L and pEG202- α 1, we amplified the cDNA encoding the intracellular loops of GABA_A-receptor γ 2S, γ 2L and α 1 subunits by the polymerase chain reaction (PCR) and inserted the amplified sequences into pEG202 between *EcoRI* and *BamHI* sites (for γ 2S and γ 2L subunits) or between *EcoRI* and *NcoI* sites (for the α 1 subunit). cDNA encoding amino acids H318 to D403 of γ 2S, H318 to D411 of γ 2L and N307 to D392 of α 1 were fused in-frame to cDNA encoding the C terminus of LexA (amino acids 1–202, including the DNA-binding domain and the dimerization domain). Deletion constructs of the γ 2L intracellular loop were generated by PCR and cloned into pEG202. Yeast co-transformed with bait plasmid and the reporter plasmid pJK103 were used for the activation assay. Yeast containing bait plasmid and the reporter plasmid pJK101 were tested by the repression assay. A human fetal brain cDNA library was cloned into pJG4-5 and expressed as fusion proteins with the SV40 nuclear-localization signal, a haemagglutinin-epitope tag and the B42 transcription-activating domain at their N termini. The library was co-transformed with bait plasmid and pJK103 into yeast strain EGY48 and screened by *LacZ* and *LEU2* expression. Plasmids were retrieved from the positive clones and sequenced. The specificity was tested by co-expressing the positive prey plasmid with different bait plasmids. Plasmid pRFHM1 expressing a transcriptionally inert fragment of the *Drosophila Bicoid* product was used as a control when the specificity of interaction was tested.

***In vitro* binding assay.** cDNAs encoding full-length GABARAP, GABARAP(36–117), GABARAP(1–68), and the intracellular loop of GABA_A-receptor γ 2L and α 1 subunits were cloned in pGEX-2T (Pharmacia) and expressed in *Escherichia coli* strain DH5 α or BL21 and purified by affinity chromatography. To produce the ³⁵S-labelled proteins using the TNT^R-coupled reticulocyte lysate system (Promega), we cloned cDNAs for the γ 2S and γ 2L intracellular loops and for GABARAP(36–117) into pETNB, a modified pET plasmid. We incubated 10 μ l of the *in vitro*-translated product with 20 μ l glutathione-agarose charged with 2 μ g GST-fusion proteins in 400 μ l 20 mM Tris, pH 7.5, 25 mM NaCl, 60 mM MgCl₂, 1 mM EDTA, 1 mM dithiothreitol (DTT), 10% glycerol, 0.5% NP40 and 1% non-fat milk (buffer A) (4°C, 1 h). The agarose was washed four times with 500 μ l buffer A less 1% non-fat milk. The bound proteins were then eluted by 5 mM glutathione/50 mM Tris, pH 8.0. The eluted proteins were resolved on SDS-PAGE and visualized by autoradiography. To test the association of tubulin with GABARAP, 50 μ g purified tubulin (from Cytoskeleton) was incubated with GST-fusion proteins. Buffer B (20 mM Tris, pH 7.5, 5 mM MgCl₂, 1 mM EDTA, 1 mM DTT, 10% glycerol, 0.5% NP40, 1% non-fat milk) was used as incubation and washing solution. The bound tubulin was detected by immunostaining with monoclonal anti- β -tubulin antibody (1:250 dilution, Sigma) and visualized by enhanced chemiluminescence (Amersham).

Generation of anti-GABARAP antibodies and immunoprecipitation. Rabbit polyclonal antisera were generated against GST-GABARAP or GABARAP and purified by Affi-Gel 10 (Bio-Rad) charged with the antigen. Membranes prepared from rat brain were solubilized in 20 mM CHAPS, 50 mM Tris, pH 8.0, 50 mM KCl, 5 mM MgCl₂, 1 mM EDTA, 2 mM phenylmethylsulphonyl fluoride and additional protease inhibitors, and incubated with antibodies against GST-GABARAP or GABARAP charged to Affi-Gel 10. The precipitates were washed four times with the same buffer. The bound proteins were eluted by boiling in SDS-PAGE loading buffer and analysed by western blotting.

Cell biology and immunocytochemistry. Rat cortical cells at embryonic day 16 were grown in medium needed to suppress the growth of glia²¹. The cultured cells were washed with PBS and fixed in 4% paraformaldehyde and 4% sucrose in PBS, permeabilized in 0.25% Triton-X-100, blocked by 5% normal goat

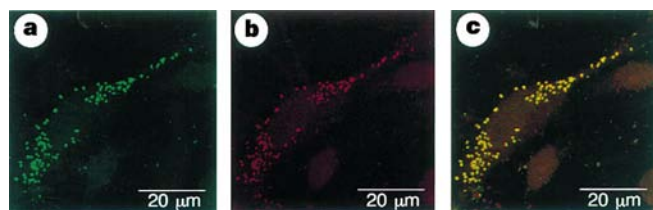


Figure 3 Co-localization of GABA_A receptors and GABARAP in cultured cortical neurons. The distribution of **a**, GABARAP and **b**, GABA_A receptors in cultured cortical neurons was shown by double immunostaining with anti-GABARAP and bd17 antibodies. GABARAP was visualized by FITC-conjugated secondary antibody (green) and GABA_A receptors by Texas red (red). **c**, Superimposition of the images in **a**, **b**. The figure is a composite of multiple focal planes, each of which showed co-localization of GABARAP and GABA_A receptors. No staining was observed when **a**, anti-GABARAP antibody was pre-absorbed by antigen-affinity column, or **b**, antibody bd17 was eliminated (data not shown).

serum, and incubated with rabbit anti-GABARAP antibody (1:300 dilution) and bd17 (5–10 $\mu\text{g ml}^{-1}$, a mouse monoclonal antibody against the $\beta 2/\beta 3$ subunit of the GABA_A receptor). After overnight incubation at 4 °C, the cells were washed with PBS and incubated with fluorescein isothiocyanate (FITC)-conjugated anti-rabbit and Texas-red-conjugated anti-mouse antibodies.

Received 24 July; accepted 4 November 1998.

- Macdonald, R. L. & Olsen, R. W. GABA_A receptor channels. *Annu. Rev. Neurosci.* **17**, 569–602 (1994).
- Leidenheimer, N. J., Browning, M. D. & Harris, R. A. GABA_A receptor phosphorylation: multiple sites, actions and artifacts. *Trends Pharmacol. Sci.* **12**, 84–87 (1991).
- Moss, S. J., Smart, T. G., Blackstone, C. D. & Huganir, R. L. Functional modulation of GABA_A receptors by cAMP-dependent protein phosphorylation. *Science* **257**, 661–665 (1992).
- Nusser, Z., Roberts, J. D., Baude, A., Richards, J. G. & Somogyi, P. Relative densities of synaptic and extrasynaptic GABA_A receptors on cerebellar granule cells as determined by a quantitative immunogold method. *J. Neurosci.* **15**, 2948–2960 (1995).
- Froehner, S. C. Regulation of ion channel distribution at synapses. *Annu. Rev. Neurosci.* **16**, 347–368 (1993).
- Carr, C., McCourt, D. & Cohen, J. B. The 43-kilodalton protein of *Torpedo* nicotinic postsynaptic membranes: purification and determination of primary structure. *Biochemistry* **26**, 7090–7102 (1987).
- Prior, P. et al. Primary structure and alternative splice variants of gephyrin, a putative glycine receptor-tubulin linker protein. *Neuron* **8**, 1161–1170 (1992).
- Zervos, A. S., Gyuris, J. & Brent, R. Mxi1, a protein that specifically interacts with Max to bind Myc-Max recognition sites. *Cell* **72**, 223–232 (1993).
- Fritschy, J. M. & Mohler, H. GABA_A-receptor heterogeneity in the adult rat brain: differential regional and cellular distribution of seven major subunits. *J. Comp. Neurol.* **359**, 154–194 (1995).
- Moss, S. J., Doherty, C. A. & Huganir, R. L. Identification of the cAMP-dependent protein kinase and protein kinase C phosphorylation sites within the major intracellular domains of $\beta 1$, $\gamma 2S$ and $\gamma 2L$ subunits of γ -aminobutyric acid type A receptor. *J. Biol. Chem.* **267**, 14470–14476 (1992).
- Mann, S. S. & Hammarback, J. A. Molecular characterization of light chain 3. *J. Biol. Chem.* **269**, 11492–11497 (1994).
- Maccioni, R. B. & Cambiasso, V. Role of microtubule-associated proteins in the control of microtubule assembly. *Physiol. Rev.* **75**, 835–864 (1995).
- Sanchez, C., Padilla, R., Paciucci, R., Zabala, J. C. & Avila, J. Binding of heat-shock protein 70 (hsp70) to tubulin. *Arch. Biochem. Biophys.* **310**, 428–432 (1994).
- Garnier, J., Osguthorpe, D. J. & Robson, B. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.* **120**, 97–120 (1978).
- Gao, B. & Fritschy, J. M. Selective allocation of GABA_A receptors containing the $\alpha 1$ subunit to neurochemically distinct subpopulations of rat hippocampal interneurons. *Eur. J. Neurosci.* **6**, 837–853 (1994).
- Hill, J. A. Jr Nicotinic receptor-associated 43K protein and progressive stabilization of the postsynaptic membrane. *Mol. Neurobiol.* **6**, 1–17 (1992).
- Langosch, D., Hoch, W. & Betz, H. The 93 kDa protein gephyrin and tubulin associated with the inhibitory glycine receptor are phosphorylated by an endogenous protein kinase. *FEBS Lett.* **298**, 113–117 (1992).
- Gillespie, K. H., Balasubramanian, S., Fung, E. T. & Huganir, R. L. Rapsyn clusters and activates the synapse-specific receptor tyrosine kinase MuSK. *Neuron* **16**, 953–962 (1996).
- Kannenbergh, K., Baur, R. & Sigel, E. Proteins associated with alpha 1-subunit-containing GABA_A receptors from bovine brain. *J. Neurochem.* **68**, 1352–1360 (1997).
- Item, C. & Sieghart, W. Binding of gamma-aminobutyric acid A receptors to tubulin. *J. Neurochem.* **63**, 1119–1125 (1994).
- Nishimura, R. N. & Dwyer, B. E. Evidence for different mechanisms of induction of HSP70: a comparison of cultured rat cortical neurons with astrocytes. *Mol. Brain Res.* **36**, 227–239 (1996).

Acknowledgements. We thank R. Brent for the yeast two-hybrid system; K. Wilcox for pETNB plasmid; Z. Nusser and C. Houser for helpful discussion; and J. DeVellis and R. Cole for help with primary neuronal culture. This work is supported by an NIH grant (to R.W.O.) and an NIH training grant (to H.W.).

Correspondence and requests for materials should be addressed to R.W.O. (e-mail: rolsen@mednet.ucla.edu).

The AMPA receptor interacts with and signals through the protein tyrosine kinase Lyn

Takashi Hayashi*, Hisashi Umemori*, Masayoshi Mishina† & Tadashi Yamamoto*

* Department of Oncology, Institute of Medical Science, The University of Tokyo, 108-8639 Tokyo, Japan

† Department of Molecular Neurobiology and Pharmacology, School of Medicine, The University of Tokyo, 113-0033 Tokyo, and CREST, Japan Science and Technology Corporation, 332-0012 Saitama, Japan

Glutamate is the major excitatory neurotransmitter in the mammalian central nervous system. The ionotropic glutamate receptors are classified into two groups, NMDA (N-methyl-D-aspartate) receptors and AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionate) receptors. The AMPA receptor is a ligand-gated cation channel that mediates the fast component of excitatory

postsynaptic currents in the central nervous system^{1,2}. Here we report that AMPA receptors function not only as ion channels but also as cell-surface signal transducers by means of their interaction with the Src-family non-receptor protein tyrosine kinase Lyn. In the cerebellum, Lyn is physically associated with the AMPA receptor and is rapidly activated following stimulation of the receptor. Activation of Lyn is independent of Ca^{2+} and Na^{+} influx through AMPA receptors. As a result of activation of Lyn, the mitogen-activated protein kinase (MAPK) signalling pathway is activated, and the expression of brain-derived neurotrophic factor (BDNF) messenger RNA is increased in a Lyn-kinase-dependent manner. Thus, AMPA receptors generate intracellular signals from the cell surface to the nucleus through the Lyn–MAPK pathway, which may contribute to synaptic plasticity by regulating the expression of BDNF.

Among the Src-family protein tyrosine kinases (PTKs), Src, Yes, Fyn and Lyn show high expression in the central nervous system (CNS)^{3–8}. As several ion channels are physically associated with Src^{9,10}, we investigated whether the AMPA receptor interacts with Src-family PTKs. We generated antibodies that recognize the ionotropic glutamate receptors GluR2 and GluR3 (anti-GluR2/3)¹¹. The AMPA receptor was co-immunoprecipitated with Lyn (Fig. 1a, b), suggesting their physical association in the cerebellum. We detected the AMPA receptor at very low levels in the anti-Src, anti-Yes and anti-Fyn immunoprecipitates. Conversely, Lyn was also co-immunoprecipitated with the AMPA receptor (Fig. 1c). Similar results were obtained using antibodies that recognize other subunits of the AMPA receptor (GluR1, GluR3 and GluR4) (data not shown). Because the AMPA receptor is heteromeric and assembled from four related subunits¹, it is not known whether Lyn associates with every subunit or with a particular subunit.

We estimated that about 1–2% of Lyn associates with the AMPA receptor in the cerebellum (data not shown). To identify the domain of Lyn that mediates its binding to the AMPA receptor, we generated a series of glutathione S-transferase (GST) fusion proteins containing specific domains of Lyn. Binding experiments showed that the Src homology (SH) region is involved in the association of Lyn with the AMPA receptor (Fig. 1d). The three-dimensional structure of Src-family PTKs showed that, in the inactive protein, the SH3 region binds the linker region that connects the SH2 and catalytic domain^{12,13}; this linker forms a left-handed polyproline type II helix, as seen in peptides having the consensus sequence PXXPR. Because the AMPA receptor is constitutively associated with Lyn (see below) and because the AMPA receptor does not carry a proline-rich sequence, a sequence outside the polyproline-recognition regions in the Lyn SH3 domain would be involved in binding to the AMPA receptor.

Co-immunoprecipitation experiments indicated that Lyn associated physically with the AMPA receptor in HEK 293T cells expressing the mouse AMPA-receptor subunit GluR2² and Lyn (Fig. 1e). To localize the sites on the AMPA receptor that interact with Lyn, we generated GST-fusion proteins containing an intracellular region of GluR2. Using these GST–GluR2 fusion proteins, we determined that the intracellular region between the fourth transmembrane domain and the carboxyl terminus is responsible for binding of GluR2 to Lyn (Fig. 1f).

We observed the activation of Lyn PTK upon stimulation of AMPA receptors (with 100 μM AMPA) in cerebellar primary culture cells in the absence of extracellular Ca^{2+} (Fig. 2a). In this and subsequent experiments, cells were pretreated with the sodium-channel blocker tetrodotoxin (TTX) to reduce a secondary release of endogenous neurotransmitters from cultured neurons. Stimulation with low concentrations of AMPA (1 μM) or with NMDA (500 μM) failed to increase the level of autophosphorylation of Lyn in cerebellar primary culture cells (Fig. 2b), indicating that Lyn was probably not activated by these stimuli. Therefore, stimulation of

AMPA receptors with a high concentration of its agonist may be required to activate the signal-transducer function of the AMPA receptor; that is, Lyn would be activated only following robust synaptic activation. The effect of AMPA was suppressed by the selective AMPA-receptor antagonist GYKI 52466 and enhanced by cyclothiazide, which blocks desensitization of AMPA receptors and also slows unbinding of the agonist (Fig. 2c). The extent of physical association of the AMPA receptor with Lyn, however, was unaltered by stimulation of the AMPA receptor (Fig. 2d).

Stimulation of AMPA receptors resulted in a rapid (within 1 minute) and significant increase in the PTK activity of Lyn in HEK 293T cells expressing exogenous GluR2 and Lyn: 2.9 ± 0.7 (mean $\pm$ s.e.m.)-fold stimulation of Lyn autophosphorylation and 2.5 ± 0.7 -fold stimulation of enolase phosphorylation occurred within 3 minutes of AMPA-receptor stimulation ($n = 6$; typical data are shown in Fig. 2e, left). Activation of the Lyn PTK following stimulation with glutamate was not observed in HEK 293T cells transfected with Lyn only (Fig. 2e, right). Thus, the GluR2 subunit of AMPA receptors was sufficient to activate the Lyn PTK following receptor stimulation. AMPA receptors consisting of only the GluR2 subunit cannot function as ion channels^{1,2}; thus, channel activity of the AMPA receptor is apparently not required for

activation of Lyn. We therefore assume that a conformational change in the AMPA receptor is rapidly induced upon the stimulation of the receptor and that the conformational change is relevant to the activation of the Lyn PTK activity.

As PTKs link to the Ras/MAPK signalling pathway, activation of Lyn by AMPA-receptor stimulation may lead to activation of MAPK in neurons. MAPK is activated by non-NMDA-receptor-mediated activity in primary culture neurons¹⁴. We therefore determined whether Lyn was involved in AMPA-receptor-mediated activation of MAPK in cerebellar primary culture cells. MAPK was activated upon addition of AMPA (100 μ M) to cerebellar primary culture cells, and MAPK activation through AMPA receptors was inhibited by the PTK inhibitors genistein or herbimycin A (Fig. 3a). To test for direct involvement of Lyn PTK activity in AMPA-receptor-mediated MAPK activation, we generated recombinant adenoviruses carrying wild-type Lyn (AdexLyn) or the kinase-negative form of Lyn (AdexLynKN). The recombinant replication-deficient type 5 adenovirus (Adex) enables direct gene transfer into the neurons, and high titres of Adex infected nearly all cerebellar primary culture cells, as described previously^{15,16}. MAPK activation in response to AMPA-receptor stimulation was enhanced by infection with Adex-Lyn. Moreover, MAPK activation was almost completely inhibited

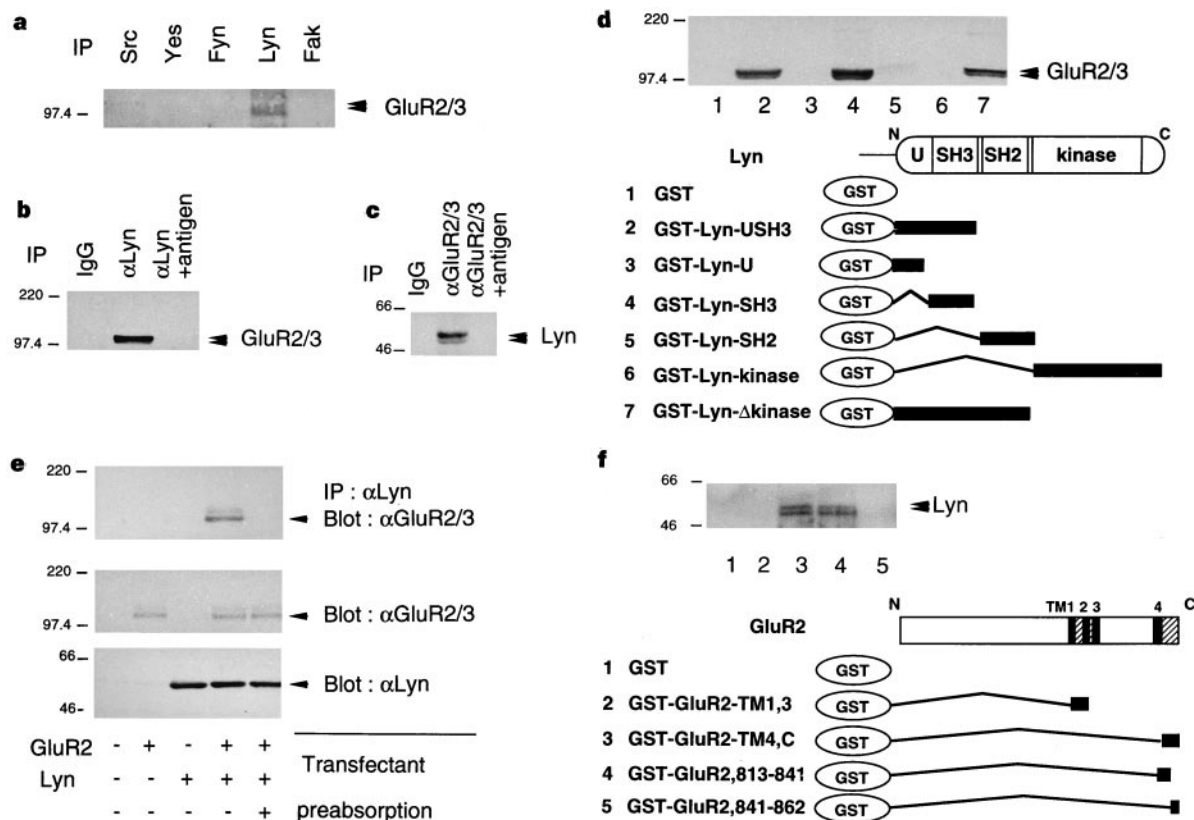


Figure 1 Association of the AMPA receptor with Lyn. **a**, The AMPA receptor GluR2/3 subunit in immunoprecipitates (IP) produced using anti-PTK antibodies (anti-Src, anti-Yes, anti-Fyn, anti-Lyn, anti-Fak) in mouse cerebellar lysates. **b**, The AMPA receptor in immunoprecipitates produced using anti-Lyn antibody (α Lyn), anti-Lyn antibody preabsorbed with antigen²⁶ or nonspecific IgG in mouse cerebellar lysates. **c**, Lyn PTK in immunoprecipitates produced using anti-AMPA receptor (GluR2/3) antibodies, anti-AMPA receptor (GluR2/3) antibodies preabsorbed with antigen¹¹ or nonspecific rabbit IgG in mouse cerebellar lysates. **d**, The AMPA receptor in precipitates from mouse cerebellar lysates produced using GST-Lyn fusion proteins. Proteins used were GST alone (lane 1) or GST-fusion proteins containing specific domains of Lyn: the unique and SH3 domains (GST-Lyn-USH3, lane 2), the unique domain (GST-Lyn-U, lane 3), the SH3 domain (GST-Lyn-SH3, lane 4), the SH2 domain (GST-Lyn-SH2, lane 5) and the catalytic domain (GST-Lyn-kinase, lane 6). GST-Lyn- Δ kinase (lane 7) is a GST-Lyn fusion protein that lacks the Lyn catalytic domain. **e**, GluR2 in anti-Lyn immunoprecipitates or anti-Lyn immunoprecipitates preabsorbed with antigen²⁶ from the lysates of HEK 293T cells transfected with plasmids encoding GluR2 and/or plasmids encoding Lyn PTK (top). Expression of each protein was confirmed by protein immunoblots of cell lysates probed with antibodies to the respective proteins (middle and bottom). Similar results were obtained using COS cells (data not shown). **f**, Lyn in the precipitates from mouse cerebellar lysates produced using GST-GluR2 fusion proteins. Proteins used were GST alone (lane 1) or GST-fusion proteins containing intracellular regions of GluR2: the region between transmembrane domains 1 and 3 (GST-GluR2-TM1,3, lane 2) and the region between transmembrane domain 4 and the C terminus (GST-GluR2-TM4,C, lane 3; GST-GluR2,813-841, lane 4; GST-GluR2,841-862, lane 5). Relative molecular masses (in K) of markers are shown at the left. Data are representative of: **a**, seven, **b**, three, **c**, six, **d**, five, **e**, nine and **f**, three experiments.

protein that lacks the Lyn catalytic domain. **e**, GluR2 in anti-Lyn immunoprecipitates or anti-Lyn immunoprecipitates preabsorbed with antigen²⁶ from the lysates of HEK 293T cells transfected with plasmids encoding GluR2 and/or plasmids encoding Lyn PTK (top). Expression of each protein was confirmed by protein immunoblots of cell lysates probed with antibodies to the respective proteins (middle and bottom). Similar results were obtained using COS cells (data not shown). **f**, Lyn in the precipitates from mouse cerebellar lysates produced using GST-GluR2 fusion proteins. Proteins used were GST alone (lane 1) or GST-fusion proteins containing intracellular regions of GluR2: the region between transmembrane domains 1 and 3 (GST-GluR2-TM1,3, lane 2) and the region between transmembrane domain 4 and the C terminus (GST-GluR2-TM4,C, lane 3; GST-GluR2,813-841, lane 4; GST-GluR2,841-862, lane 5). Relative molecular masses (in K) of markers are shown at the left. Data are representative of: **a**, seven, **b**, three, **c**, six, **d**, five, **e**, nine and **f**, three experiments.

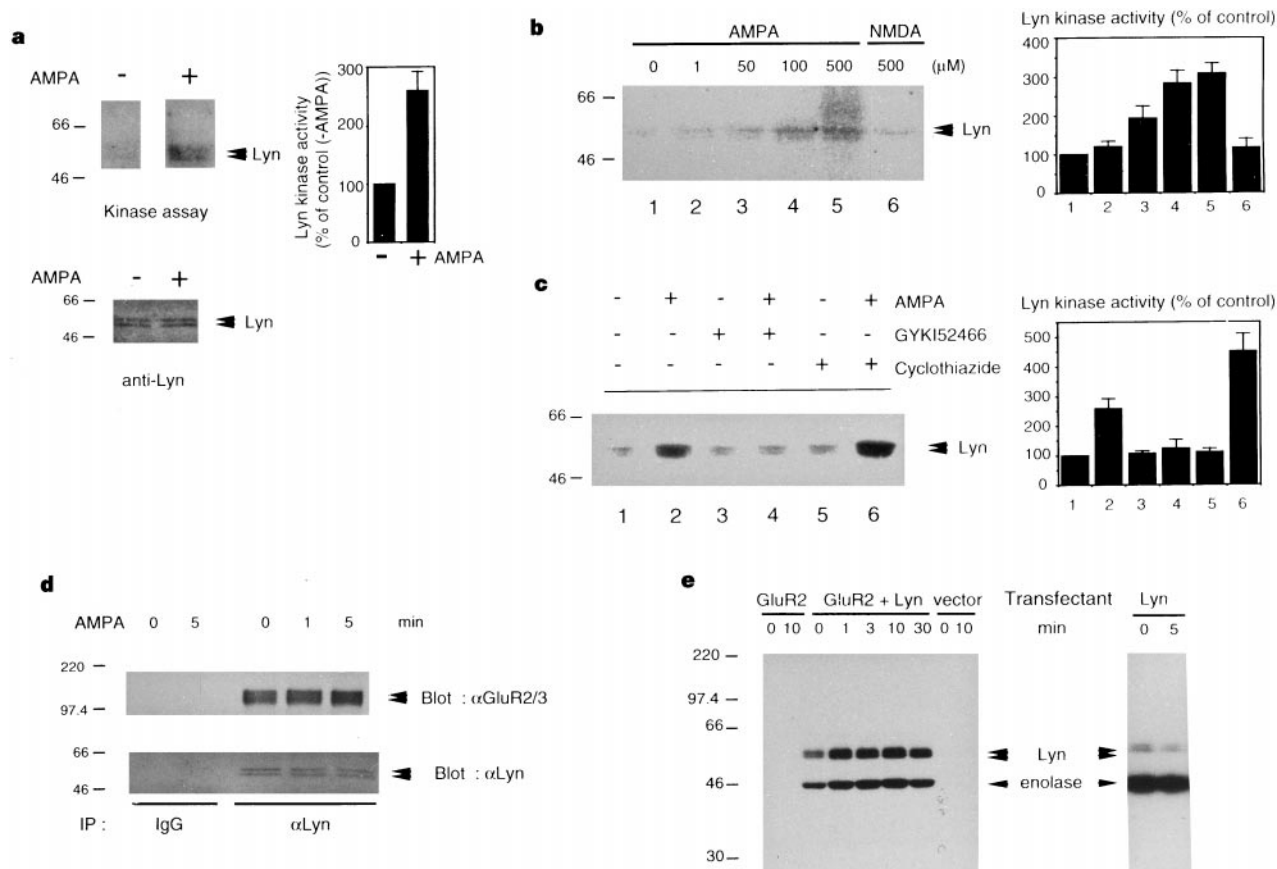


Figure 2 Activation of Lyn following stimulation of AMPA receptors. **a**, Cerebellar primary culture cells were stimulated (+) or not stimulated (-) with 100 μ M AMPA for 4 min. Top, autophosphorylation of Lyn in the kinase assay. Bottom, western blot analysis of Lyn in cerebellar primary culture cells. **b**, Concentration-dependent effects of AMPA and NMDA on Lyn PTK activation. **c**, Blockade of the effect of AMPA on Lyn by 100 μ M GYKI 52466 and enhancement of this effect by 100 μ M cyclothiazide added 3 min before 100 μ M AMPA. **d**, The AMPA receptor and Lyn proteins in the anti-Lyn immunoprecipitates or pre-immune

control immunoprecipitates (IgG). The lysates for immunoprecipitations were prepared from mouse cerebellar primary culture cells following stimulation with 100 μ M AMPA for the indicated time periods. **e**, HEK 293T cells transfected with GluR2 and two forms of Lyn were stimulated with 500 μ M glutamate for the indicated time periods. Both autophosphorylation of Lyn and phosphorylation of exogenous substrate, enolase, are shown. Relative molecular masses (in K) of markers are shown at the left. Data are representative of: **a**, six; **b**, four; **c**, four; **d**, six and **e**, six experiments.

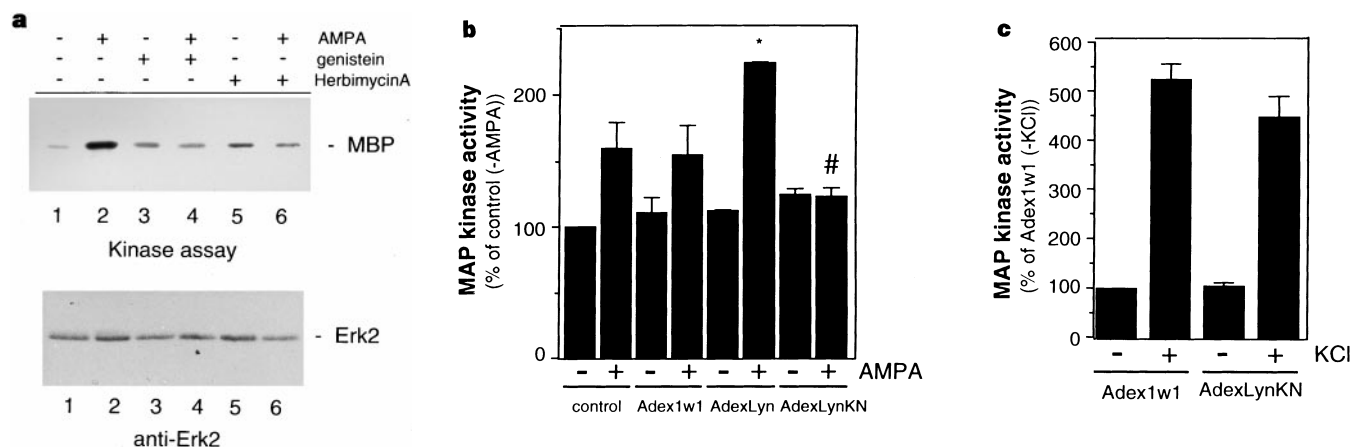


Figure 3 Activation of the MAPK pathway in cerebellar primary culture cells. Activation of MAPK after stimulation with **a**, **b**, 100 μ M AMPA or **c**, 50 mM KCl for 5 min. **a**, the effects of genistein (lanes 3, 4) and herbimycin A (lanes 5, 6) on the activation of the MAPK Erk2 in culture cells stimulated by AMPA (lanes 2, 4, 6) or unstimulated (lanes 1, 3, 5). Cerebellar primary culture cells were treated with 100 μ M genistein or 10 μ M herbimycin A before being stimulated with AMPA (kinase assay, top). Myelin basic protein (MBP) was used as a MAPK substrate.

Bottom, western blot analysis of Erk2 in cerebellar primary culture cells. **b**, **c**, Kinase activities of Erk2 from the lysates of cerebellar primary culture cells infected with Adex viruses shown below the columns (multiplicity of infection 10–20). Adex1w1 is a control virus that lacks expressing units¹⁵. The results are the means $\pm$ s.e.m., $n = 4$; asterisk indicates $P < 0.01$; hash symbol indicates $P < 0.01$.

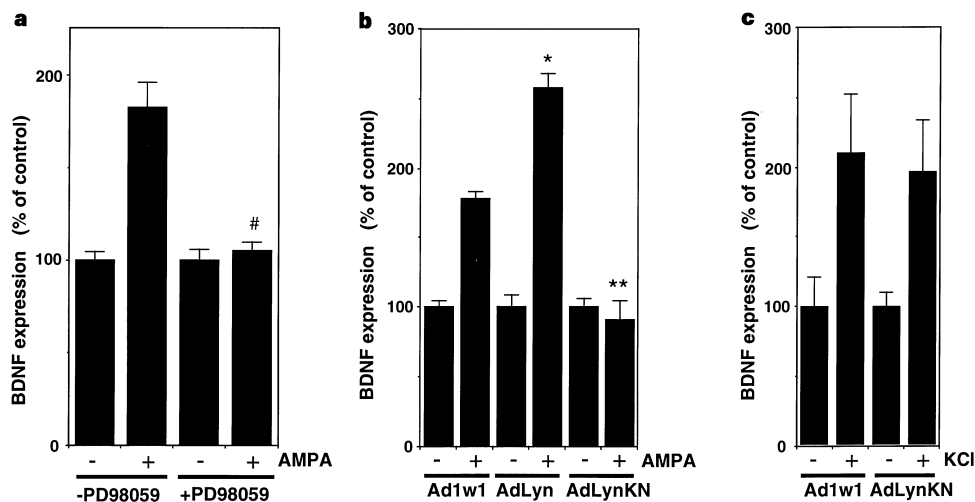


Figure 4 BDNF gene expression in cerebellar primary culture cells. BDNF expression in cells stimulated with **a**, **b**, 50 μ M AMPA or **c**, 50 mM KCl for 4 h. Ratios of BDNF mRNA to GAPDH mRNA (the control) are shown. **a**, The blockade of AMPA-receptor-regulated BDNF gene expression by pretreatment with the MEK inhibitor PD98059 (30 μ M). **b**, **c**, Effects of wild-type or kinase-negative forms

of Lyn on BDNF expression. Infected Adex viruses are shown below the columns (multiplicity of infection 10–20). The results are the means $\pm$ s.e.m., **a**, $n = 4$, **b**, $n = 6$, **c**, $n = 4$; hash symbol indicates $P < 0.01$; asterisk indicates $P < 0.001$; double asterisk indicates $P < 0.001$.

when cerebellar primary culture cells were infected with AdexLynKN (Fig. 3b). MAPK could be activated by other pathways, such as K^+ -induced depolarization, in cells infected with AdexLynKN (Fig. 3c). Infection with Adex had no obvious cytotoxic effects on the cells (data not shown). Thus, the Lyn PTK is involved in the AMPA-receptor-mediated activation of MAPK signalling pathways. We conclude that Lyn is physically and functionally associated with the AMPA receptor in the cerebellum.

Activated MAPK translocates into the nucleus and may regulate gene expression¹⁷. Neurotrophic factors, such as BDNF, are involved in both developmental and adult synaptic plasticity^{18,19}. The synthesis of neurotrophins is regulated by neuronal activity mediated by non-NMDA receptors^{20,21}. However, little is known about the intracellular signalling pathways that mediate induction of neurotrophins. To determine the biological significance of the AMPA-receptor-mediated activation of MAPK signalling pathways, we studied the role of MAPK in AMPA-receptor-mediated induction of BDNF gene expression. PD98059, an inhibitor of MAPK–Erk kinase (MEK), which blocks MAPK activation, reduced AMPA-receptor-induced BDNF gene expression in cerebellar primary culture cells (Fig. 4a). Thus, the MAPK signalling pathway transmits the signal from the AMPA receptor to activate BDNF gene expression. Moreover, BDNF gene expression in response to AMPA-receptor stimulation was enhanced by infection with AdexLyn and was reduced by infection with AdexLynKN (Fig. 4b). BDNF gene expression was not affected by AdexLynKN when cerebellar primary cultured cells were depolarized by a Lyn-independent pathway, such as a high concentration of K^+ (Fig. 4c). Thus, BDNF gene expression is regulated by Lyn-mediated activation of MAPK in response to AMPA-receptor signalling. Expression of BDNF in our experiments (in which neurons were cultured for 3–4 weeks) was lower than that in previous studies, in which younger culture neurons (cultured for <1 week)^{20,21} were used. This difference in expression might be attributable to the difference between the states of the cultured neurons in the developing periods and the states that in the periods after the neuronal network connections have been completed.

We have shown that the AMPA receptor on the postsynaptic membrane relays signals to the nucleus by means of its interaction with the Src-family PTK Lyn. Our results indicate that the AMPA receptor acts not only as a ligand-gated ion channel in normal synaptic transmission, but also as a sort of postsynaptic sensor that recognizes only the strong stimulation from the presynaptic to the

postsynaptic membrane. The idea that the AMPA receptor functions as a signal transducer has been introduced previously²⁸. We have further showed that the Lyn-mediated signalling pathway changes the expression of BDNF. BDNF strengthens the efficiency of synaptic transmission by refining stimulated synapses in the CNS^{18,19}. Long-lasting synaptic plasticity may require PTK activity^{6,22–24} and the expression of the neurotrophins BDNF and neurotrophin-3 (refs 18, 19, 24). Thus, Lyn-mediated signalling through AMPA receptors may contribute to transforming short-term plasticity to long-term plasticity, a process that involves gene expression. This signalling pathway may be important for the molecular and cellular basis of synaptic plasticity in the cerebellum. □

Methods

Cell culture, transfection and stimulation. A primary culture from the cerebellum was prepared from fetal Wistar rats or ICR mice (postnatal day 0 (P0) to P1) as described²⁵. Cultured neurons (10^6 cells per lane) from between days 19 and 28 were used. HEK 293T cells and COS cells (10^7 cells per lane) were transfected with pME18S-Lyn (SR α promoter) and/or pcDNA3-GluR2 (CMV promoter) by the calcium phosphate method as described⁴. For stimulation of AMPA receptors, cultures were washed twice with 5 ml extracellular solution containing, in mM: HEPES, 15 (pH 7.5); NaCl, 140; KCl, 5; MgCl₂, 1; glucose, 5; 0.12% NaHCO₃ (and CaCl₂, 2 in Fig. 3c). 5 μ M TTX was added to inhibit neuronal activity. Then cultures were incubated in 4 ml of the same buffer at 37 °C for 30 min before stimulation with GluR agonists.

Immunoprecipitation and immunoblot. Rodent brain or HEK 293T or COS cells expressing AMPA receptors, and PTKs were lysed in TNE buffer containing, in mM: Tris-HCl, 50 (pH 8.0); 1% (w/v) NP40; EDTA, 20; sodium orthovanadate, 1; NaF, 1; 10 μ g ml⁻¹ leupeptin; 10 μ g ml⁻¹ aprotinin). Proteins in the lysates were immunoprecipitated with antibodies (described below) and then incubated with protein A- or protein G-Sepharose (Figs 1a, b, c, e, 2, 3), or were precipitated with GST-fusion proteins (10 μ g protein per ml lysate) and then incubated with glutathione-Sepharose beads (Fig. 1d). GST–Lyn–USH3, GST–Lyn–U, GST–Lyn–SH3, GST–Lyn–SH2, GST–Lyn–kinase and GST–Lyn– Δ kinase encoded Lyn amino acids 1–118, 1–57, 58–118, 119–235, 236–512 and 1–235, respectively. GST–GluR2–TM1,3, GST–GluR2–TM4,C, GST–GluR2,813–841 and GST–GluR2,841–862 encoded amino acids 545–599, 813–862, 813–841 and 841–862 of GluR2, respectively. Samples were then subjected to SDS–PAGE followed by immunoblotting. The monoclonal antibodies GD11 (against Src), 3H9 (against Yes), Lyn-8 and Lyn-9 (against Lyn), Fyn-301 (against Fyn) and 2A7 (against Fak) have been described previously^{4,5,26,27}. Polyclonal antibodies to each AMPA-receptor subunit were

obtained from rabbits immunized with synthetic peptides corresponding to amino-acid sequences in the C terminus of each subunit¹¹. Anti-Erk2 antibodies were purchased from Santa Cruz.

Kinase assay. Immune complexes containing Lyn or MAPK (Erk2) were prepared from cell lysates and used in the *in vitro* kinase reaction^{4,26,27}. Cultured cells were lysed in RIPA buffer containing, in mM: Tris-HCl, 10 (pH 7.5); 1% (w/v) NP40; 0.1% (w/v) sodium deoxycholate; 0.1% (w/v) SDS; NaCl, 150; EDTA, 1; sodium orthovanadate, 1; NaF, 1; 10 $\mu\text{g ml}^{-1}$ leupeptin; 10 $\mu\text{g ml}^{-1}$ aprotinin (Figs 2, 3). Lyn PTK was assayed *in vitro* using enolase as a substrate in reaction buffer containing, in mM: HEPES, 20 (pH 7.5); MgCl₂, 10; MnCl₂, 5; 10 μM ATP; 1 μCi [γ -³²P]ATP. MAP kinase was assayed *in vitro* using myelin basic protein as a substrate in reaction buffer containing, in mM: Tris-HCl, 20 (pH 8.0); MgCl₂, 10; MnCl₂, 5; 25 μM [γ -³²P]ATP.

Northern analysis. Total RNA was prepared by the acid guanidinium-phenol-chloroform method^{20,21} from cerebellar primary culture cells stimulated or unstimulated with 50 μM AMPA or 50 mM KCl for 4 h in culture medium²⁵. RNA blot analysis was done with standard procedures^{20,21} and labelled probes were prepared from rat BDNF complementary DNA and GAPDH cDNA with the Megaprime DNA labelling kit (Amersham). Quantification was done with the BAS 2000 system.

Received 23 July; accepted 22 October 1998.

- Hollmann, M. & Heinemann, S. Cloned glutamate receptors. *Annu. Rev. Neurosci.* **17**, 31–108 (1994).
- Sakimura, K. *et al.* Functional expression from cloned cDNAs of glutamate receptor species responsive to kainate and quisqualate. *FEBS Lett.* **272**, 73–80 (1990).
- Wagner, K. R., Mei, L. & Huganir, R. L. Protein tyrosine kinases and phosphatases in the nervous system. *Curr. Opin. Neurobiol.* **1**, 65–73 (1991).
- Umemori, H., Sato, S., Yagi, T., Aizawa, S. & Yamamoto, T. Initial events of myelination involve Fyn tyrosine kinase signalling. *Nature* **367**, 572–576 (1994).
- Sudol, M., Alvarez-Buylla, A. & Hanafusa, H. Differential developmental expression of cellular *yes* and cellular *src* proteins in cerebellum. *Oncogene Res.* **2**, 345–355 (1988).
- Grant, S. G. N. *et al.* Impaired long-term potentiation, spatial learning, and hippocampal development in *fyn* mutant mice. *Science* **258**, 1903–1910 (1992).
- Umemori, H. *et al.* Specific expressions of Fyn and Lyn, lymphocyte antigen receptor-associated tyrosine kinases, in the central nervous system. *Mol. Brain Res.* **16**, 303–310 (1992).
- Chen, S., Ren, Y. Q. & Hillman, D. E. Transient expression of *lyn* gene in Purkinje cells during cerebellar development. *Dev. Brain Res.* **92**, 140–146 (1996).
- Holmes, T. C., Fadool, D. A., Ren, R. & Levitan, I. B. Association of Src tyrosine kinase with a human potassium channel mediated by SH3 domain. *Science* **274**, 2089–2091 (1996).
- Yu, X.-M., Askalan, R., Keil, G. J. II & Salter, M. W. NMDA channel regulation by channel-associated protein tyrosine kinase. *Science* **275**, 674–678 (1997).
- Wenthold, R. J., Yokotani, N., Doi, K. & Wada, K. Immunohistochemical characterization of the non-NMDA glutamate receptor using subunit-specific antibodies. *J. Biol. Chem.* **267**, 501–507 (1992).
- Xu, W., Harrison, S. C. & Eck, M. J. Three-dimensional structure of the tyrosine kinase c-Src. *Nature* **385**, 595–602 (1997).
- Sicheri, F., Moarefi, I. & Kuriyan, J. Crystal structure of the Src family tyrosine kinase Hck. *Nature* **385**, 602–609 (1997).
- Fiore, R. S., Murphy, T. H., Sanghera, J. S., Pelech, S. L. & Baraban, J. M. Activation of p42 mitogen-activated protein kinase by glutamate receptor stimulation in rat primary cortical cultures. *J. Neurochem.* **61**, 1626–1633 (1993).
- Miyake, S. *et al.* Efficient generation of recombinant adenoviruses using adenovirus DNA-terminal protein complex and cosmid bearing the full-length virus genome. *Proc. Natl Acad. Sci. USA* **93**, 1320–1324 (1996).
- Le Gal La Salle, G. *et al.* An adenovirus vector for gene transfer into neurons and glia in the brain. *Science* **259**, 988–990 (1993).
- Treisman, R. Regulation of transcription by MAP kinase cascades. *Curr. Opin. Cell Biol.* **8**, 205–215 (1996).
- Thoenen, H. Neurotrophins and neuronal plasticity. *Science* **270**, 593–598 (1995).
- Lo, D. C. Neurotrophic factors and synaptic plasticity. *Neuron* **15**, 979–981 (1995).
- Zafra, F., Hengeler, B., Leibrock, J., Thoenen, H. & Lindholm, D. Activity dependent regulation of BDNF and NGF mRNAs in the rat hippocampus is mediated by non-NMDA glutamate receptors. *EMBO J.* **9**, 3545–3550 (1990).
- Bessho, Y., Nakanishi, S. & Nawa, H. Glutamate receptor agonists enhance the expression of BDNF mRNA in cultured cerebellar granule cells. *Mol. Brain Res.* **18**, 201–208 (1993).
- O'Dell, T. J., Kandel, E. R. & Grant, S. G. N. Long-term potentiation in the hippocampus is blocked by tyrosine kinase inhibitors. *Nature* **353**, 558–560 (1991).
- Boxall, A. R., Lancaster, B. & Garthwaite, J. Tyrosine kinase is required for long-term depression in the cerebellum. *Neuron* **16**, 805–813 (1996).
- Kang, H. & Schuman, E. M. Long-lasting neurotrophin-induced enhancement of synaptic transmission in the adult hippocampus. *Science* **267**, 1658–1662 (1995).
- Hirano, T., Kubo, Y. & Wu, M. Cerebellar granule cells in culture: monosynaptic connections with Purkinje cells and ionic currents. *Proc. Natl Acad. Sci. USA* **83**, 4957–4961 (1986).
- Yamanashi, Y., Kakiuchi, T., Mizuguchi, J., Yamamoto, T. & Toyosima, K. Association of B cell antigen receptor with protein tyrosine kinase Lyn. *Science* **251**, 192–194 (1991).
- Yamanashi, Y. *et al.* Activation of Src-like protein-tyrosine kinase Lyn and its association with phosphatidylinositol 3-kinase upon B-cell antigen receptor-mediated signaling. *Proc. Natl Acad. Sci. USA* **89**, 1118–1122 (1992).
- Wang, Y., Small, D. L., Stanimirovic, D. B., Morley, P. & Durkin, J. P. AMPA receptor-mediated regulation of a G_i-protein in cortical neurons. *Nature* **389**, 502–504 (1997).

Acknowledgements. We thank T. Hirano, D. W. Saffen, Y. Hayashi, Y. W. Yamanashi and J. Inoue for valuable discussions; I. Saito and Y. Kanegae for preparation of recombinant adenovirus; H. Nawa for rat BDNF cDNA; and S. Nagasawa for care of mice. This work was supported by grants from the Ministry of Education, Science and Culture of Japan.

Correspondence and requests for materials should be addressed to T.Y. (e-mail: tyamamoto@ims.u-tokyo.ac.jp).

A LIM-homeodomain combinatorial code for motor-neuron pathway selection

Stefan Thor*†, Siv G. E. Andersson‡, Andrew Tomlinson‡ & John B. Thomas*

* Molecular Neurobiology Laboratory, The Salk Institute, PO Box 85800, San Diego, California 92186, USA

‡ Center for Neurobiology and Behavior, Department of Genetics and Development, College of Physicians and Surgeons of Columbia University, 701 West 168th Street, New York, New York 10032, USA

Different classes of vertebrate motor neuron that innervate distinct muscle targets express unique combinations of LIM-homeodomain transcription factors^{1,2}, suggesting that a combinatorial code of LIM-homeodomain proteins may underlie the control of motor-neuron pathway selection. Studies of LIM-homeodomain genes in mouse, *Drosophila melanogaster* and *Caenorhabditis elegans* have revealed functions of these genes in neuronal survival, axon guidance, neurotransmitter expression and neuronal function^{3–8}, but, to our knowledge, none of these studies have addressed the issue of a functional code. Here we study two members of this gene family in *Drosophila*, namely *lim3*, the homologue of the vertebrate *Lhx3* and *Lhx4* genes, and *islet*, the homologue of the vertebrate *Isl1* and *Isl2* genes. We show that *Drosophila lim3* is expressed by a specific subset of *islet*-expressing motor neurons and that mutating or misexpressing *lim3* switches motor-neuron projections predictably. Our results provide evidence that *lim3* and *islet* constitute a combinatorial code that generates distinct motor-neuron identities.

Drosophila Lim3 contains two LIM domains and a carboxy-terminal homeodomain and shows high homology to the vertebrate Lhx3/4 subclass of LIM-homeodomain proteins (Fig. 1a). The homology is highest in the homeodomain (90% identity to mouse Lhx3 and Lhx4) and is somewhat lower in the LIM domains (67%). The remainder of the sequence shows no apparent homology with Lhx3 and Lhx4 apart from a highly conserved 22-amino-acid stretch, located C-terminally to the homeodomain, that we have named the Lim3-specific domain (LSD) because it is found only in members of the Lim3 subfamily and not in other LIM-homeodomain proteins.

Using *in situ* hybridization to polytene chromosomes, we mapped the *lim3* gene to chromosome region 37B, which has been studied extensively both genetically and molecularly⁹ (Fig. 1c). The location of the *lim3* gene on a genomic walk encompassing this region indicated that *lim3* may correspond to *l(2)37Bd*, a lethal complementation group that is defined by a set of ethylmethylsulphonate-induced alleles and uncovered by several deletions¹⁰. We sequenced two strong (*Bd*¹ and *Bd*⁶) and one weak (*Bd*⁷) allele of *l(2)37Bd* and found that *Bd*¹ and *Bd*⁷ encode mutations of two different cysteine residues within the second LIM domain (Fig. 1b). Both of these cysteine residues function in the coordination of zinc ions^{11,12} and are conserved in all known LIM domains from yeast, plants and animals. The *Bd*⁶ allele has a 9-base-pair deletion in the 3' untranslated region; this deletion may affect the stability of the *Bd*⁶ messenger RNA¹³. On the basis of the precise molecular location of the *lim3* gene and the sequence analysis of these three mutant alleles, we conclude that the *l(2)37Bd* locus corresponds to the *lim3* gene.

Embryonic expression of *lim3* was confined largely to the nervous

† Present address: Department of Neurobiology, Harvard Medical School, 220 Longwood Avenue, Boston, Massachusetts 02115, USA

system, although expression was also observed in precursors of the dorsal vessel, alary muscles and in the proventriculus, as well as in a cluster of cells that correspond to either the developing ring gland or the hypocerebral ganglion (Fig. 2a, c). *lim3* transcripts were first detectable at embryonic stage 12 in subsets of cells within the brain, ventral nerve cord (VNC) and peripheral nervous system. *lim3* was not expressed by neuroblasts and appeared to be confined to neuronal progeny. In abdominal hemisegments of stage-14 embryos, *lim3* was expressed by 10–12 neurons; the number and relative positions of these cells were maintained throughout embryonic development. We were able to identify a cluster of the *lim3*-expressing cells as the RP motor neurons because of the stereotyped position of their cell bodies (Fig. 2b).

To establish the identity of the *lim3*-expressing neurons, we generated transgenic flies carrying a fusion of either *lacZ* or the *tau-myc* axon-targeted reporter⁴ with *lim3A*, a genomic fragment that directs expression of the *lacZ* or *tau-myc* reporter genes in a pattern indistinguishable from that of the *lim3* mRNA (Fig. 1c). Analysis of *lim3A-tau-myc* embryos showed that the *lim3*-expressing cells comprised a subset of interneurons and motor neurons, including the RP motor neurons (Fig. 2d–f). We detected

no Tau-Myc expression in neuroblasts, in agreement with the results obtained from *in situ* hybridization of *lim3*. In fact, most, if not all, of the Tau-Myc-positive cells appeared to extend axons shortly after the onset of Tau-Myc expression (Fig. 2d), indicating that *lim3* expression may commence in post-mitotic neurons.

The *lim3*-expressing interneurons projected axons along two discrete pathways, forming two fascicles within each longitudinal connective (Fig. 2f). To determine whether any of the *lim3*-expressing interneurons co-expressed the LIM-homeodomain-family members *apterous* (*ap*) or *islet* (*isl*), we generated embryos carrying a *lim3A* reporter transgene together with either *ap* or *isl* reporter transgenes^{4,5}. The *lim3*-expressing interneurons did not co-express either *ap* or *isl*, and neither did they project their axons in the pathways taken by either the *isl*- or the *ap*-expressing interneurons (Fig. 3).

In contrast to their non-overlapping expression pattern in interneurons, *isl* and *lim3* are co-expressed in a subset of motor neurons. The 30 somatic muscles in each *Drosophila* embryonic abdominal hemisegment are innervated by ~30 motor neurons, each of which can be uniquely identified by the position of its cell body, by its axonal projections and by innervation of its target^{14,15}. *isl* is

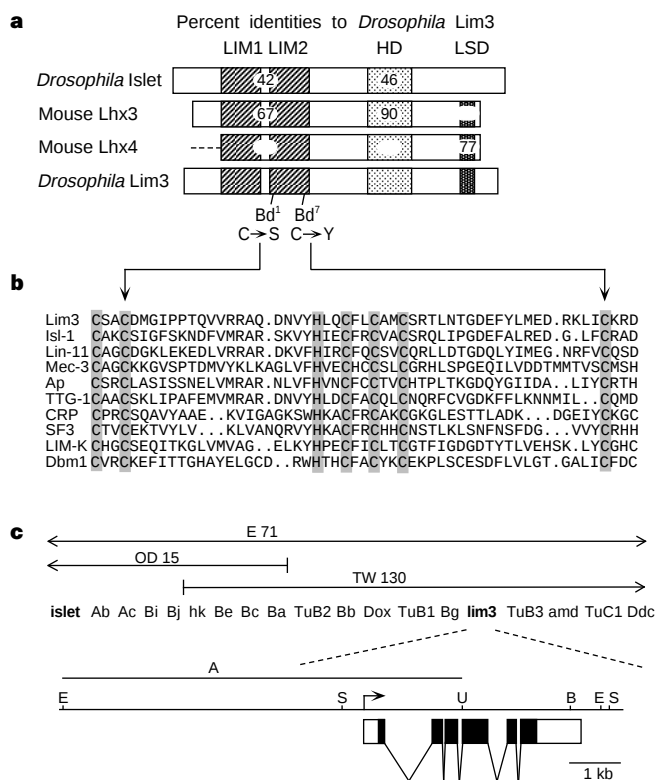


Figure 1 Molecular analysis of the *lim3* gene. **a**, The domains of *Drosophila* Lim3, aligned with other LIM-homeodomain (HD) proteins, showing the percentage of identical amino-acid residues within the homeodomain, the LIM domains and the Lim3-specific domain (LSD). The LSD consists of residues 369–392 of Lim3. **b**, Alignment of the LIM2 domains of some LIM-homeodomain proteins. The *lim3*^{Bd1} and *lim3*^{Bd7} alleles encode missense mutations of the second and sixth cysteine residues, respectively, in the LIM2 domains of Lim3. Shaded residues are conserved in all LIM domains. Species are as follows: Ap, Lim3, *Drosophila*; Isl-1, rat; Lin-11, Mec-3, *C. elegans*; TTG-1, CRP, mouse; LIM-K, human; Dbm1, *Saccharomyces cerevisiae*; SF3, sunflower. **c**, Genetic map of the region surrounding the *lim3* locus. The extent of the chromosomal deletions *Df(2L)TW130*, *Df(2L)E71* and *Df(2L)OD15* is shown above the complementation groups in the region. The genomic organization of the *lim3* gene is shown at the bottom. The arrow denotes the putative transcriptional start site. Fragment A directs expression of a reporter gene in a pattern indistinguishable from that of *lim3*. Restriction sites for *Bam*HI (B), *Eco*RI (E), *Stu*I (U) and *Sal*I (S) are indicated.

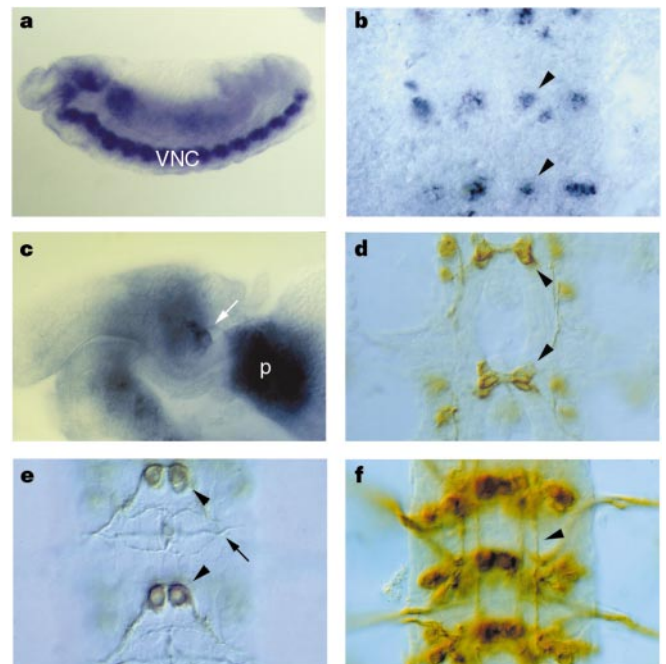


Figure 2 Embryonic expression of *lim3*. **a–c**, *In situ* hybridization analysis of *lim3* expression in stage-13, whole-mounted, wild-type embryos. **a**, Expression is seen in the developing VNC and in precursors of the developing dorsal vessel (out of the focal plane). **b**, Dorsal view of the VNC at higher magnification. By stage 13, *lim3* expression is confined to a subset of 10–12 neurons per hemisegment, including the RP motor neurons (arrowheads). **c**, *lim3* is expressed in the proventriculus (p) and in a cluster of cells adjacent to the pharynx (arrow). **d–f**, The VNC of embryos expressing *lim3A-tau-myc*; these embryos were immunostained for Tau-Myc. **d**, Tau-Myc is expressed just before the onset of axon elongation in RP motor neurons (arrowheads) in stage-13 embryos. **e**, By stage 14, the RP motor neurons (arrowheads) have extended processes across the midline and into the ISN (out of the focal plane). The arrow points to Tau-Myc-positive processes in the TN. **f**, By stage 16, the *lim3*-expressing interneurons form two axon bundles in each of the longitudinal connectives (arrowhead points to the lateral bundle).

expressed by motor neurons that project in the transverse nerve and in two different branches of the intersegmental nerve, ISNb and ISNd⁴ (Fig. 3). The ISNb motor neurons include RP1, RP3, RP4 and RP5, which innervate muscles 6, 7, 12 and 13, the contralateral V motor neuron, which innervates muscle 12, and MN28 and MN14/30, which innervate muscles 28 and 14 plus 30, respectively. The two ISNd motor neurons innervate muscles 15–17. *lim3* is co-expressed by all of the *isl*-expressing motor neurons except those projecting in ISNd. Thus, for the ISN, all of the ISNb-specific motor neurons express both *isl* and *lim3* whereas those motor neurons that project in ISNd express only *isl* (Fig. 4e). Intriguingly, a specialized VUM motor neuron that projects in both ISNb and ISNd does not express either *isl* or *lim3*.

Although *lim3*-mutant individuals die during larval and pupal stages, we did not observe any gross abnormalities in overall embryonic axonal organization in these individuals, as assayed with antibodies that recognize specific neuronal proteins, including anti-horseradish peroxidase (HRP)¹⁶ and anti-fasciclin II¹⁷ antibodies. To study the behaviour of the *lim3*-expressing neurons in detail, we crossed the *lim3A-tau-myc* transgene into *lim3* mutants. In *lim3^{Bd1}* homozygotes and hemizygotes, as well as in individuals homozygous for *TW130*, a deletion that completely removes the *lim3* locus, the *lim3*-expressing neurons survived and extended axons on schedule. However, in *lim3* mutants all of the *lim3* interneurons and motor neurons exhibited striking pathfinding defects. Here we confine our analysis to the *lim3*-expressing motor neurons of the ISN.

In wild-type embryos, the *lim3*-expressing motor neurons of the ISNb innervate the clefts between muscles 6, 7, 12, 13, 14, 28 and 30 (Fig. 4c). Using the anti-fasciclin II antibody, we found that in *lim3* mutants there is a lack of normal innervation of these muscles (Fig. 4a, b). For example, over 46% ($n = 82$) of *lim3^{Bd1}/TW130* mutant hemisegments showed no projections into the muscle-12/13 cleft, compared with 3% ($n = 67$) in wild-type hemisegments. In contrast, ISNd appeared to be thicker in *lim3^{Bd1}/TW130* mutants than in wild-type flies (Fig. 4a, b). These results indicate that ISNb motor neurons may be ectopically projecting into the ISNd target area, as would be predicted from the LIM-homeodomain code. To resolve the behaviour of the ISNb motor neurons in detail, we used the *lim3A-tau-myc* transgene to label the ISNb motor neurons

specifically (Fig. 4c, d). In wild-type embryos carrying *lim3A-tau-myc*, only 4% of hemisegments showed Tau-Myc-expressing projections into ISNd ($n = 68$). In contrast, *lim3* mutants had a high level of ectopic innervation of the ISNd target area (34%; $n = 101$ for *lim3^{Bd1}/TW130*). The numbers of these ectopic projections ranged, from segment to segment, from what appeared to be single ISNb axons to nearly all ISNb axons. Thus, mutations in *lim3* cause ISNb motor neurons to behave as ISNd motor neurons (Fig. 4f). One explanation for the variability of the change in projection is that the mutant ISNb motor neurons compete amongst themselves and with the normal ISNd motor neurons for the limited number of ISNd muscle targets, and the 'losers', by default, innervate the nearby ISNb muscles. Alternatively, the *lim3* alleles used in this study may retain some residual function, in which case null alleles might cause a more complete switch in projection.

To determine whether *lim3* is sufficient to change the identity of motor neurons, we misexpressed it using the Gal4/upstream activation sequence (UAS) system¹⁸. For these studies we used a *ftz_{ng}*-GAL4 driver, which expresses GAL4 under the control of the *fushi tarazu* (*ftz*) neurogenic enhancer¹⁹. By examining embryos that carried a single copy each of *ftz_{ng}*-GAL4.20 and a UAS-*tau-myc* transgene, we found that *ftz_{ng}*-GAL4.20 was capable of high levels of transactivation in most, if not all, motor neurons, including those projecting axons in ISNb and ISNd (Fig. 5a) (see Methods). When *lim3* was misexpressed using this GAL4 driver, many segments showed a marked increase in the number of processes in ISNb at the expense of ISNd, as assayed with the anti-fasciclin II antibody

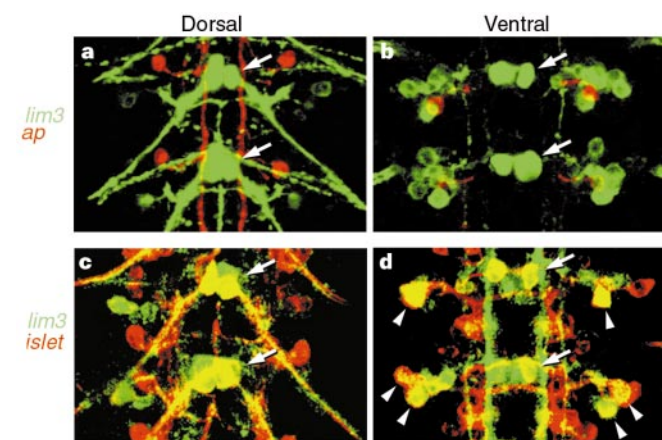


Figure 3 *lim3* expression overlaps with *isl* but not *ap* expression. **a, b**, Staining for Tau-Myc (green) and Tau-β-galactosidase (red) in a stage-16 embryo carrying the *lim3A-tau-myc* and *apC-tau-lacZ* transgenes. *lim3* and *ap* are expressed in non-overlapping sets of neurons that project axons in different pathways. Arrows point to the RP motor neurons in two adjacent segments. **c, d**, Staining for β-galactosidase (green) and Tau-Myc (red) in a stage-16 embryo carrying the *lim3A-lacZ* and *islH-tau-myc* transgenes. *lim3* is expressed by a subset of the *isl*-expressing motor neurons, including the RP motor neurons (arrows) and a lateral cluster of motor neurons that include MN14/30 and MN28 (arrowheads). Dorsal and ventral focal planes are shown.

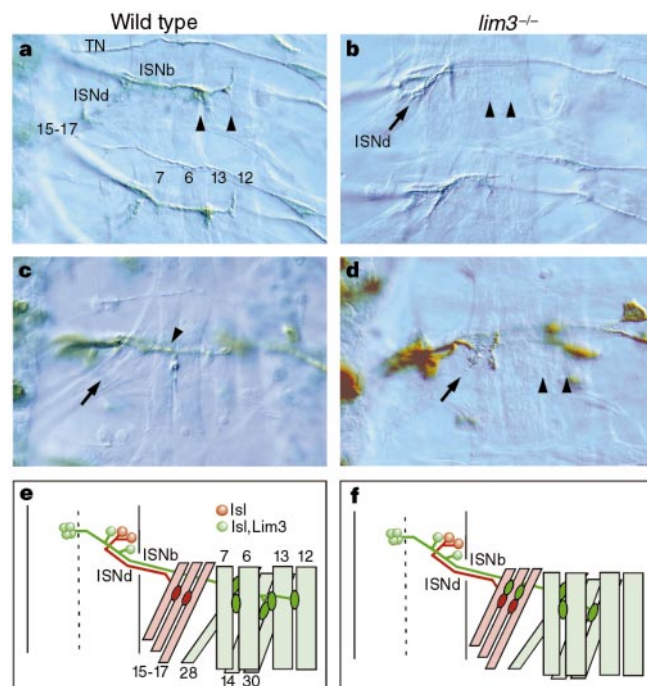


Figure 4 Defects in motor-neuron pathfinding in *lim3* mutants. **a, b**, Stage-17 embryos stained with anti-fasciclin II antibodies. **a**, In wild-type embryos, ISNb motor neurons innervate muscles 6, 7, 12 and 13 in a stereotyped fashion, extending processes into the muscle clefts (arrowheads). **b**, In *lim3^{Bd1}/TW130* mutants, the ISNb motor neurons fail to project to their specific target muscles (arrowheads) and the ISNd appears thicker than normal (arrow). **c, d**, Stage-17 embryos carrying the *lim3A-tau-myc* transgene and immunostained for Tau-Myc. **c**, In wild-type embryos, the Tau-Myc- and *lim3*-expressing motor neurons (arrowhead) project in ISNb; processes are rarely seen in ISNd (arrow). **d**, In *lim3^{Bd1}/TW130* mutants, the Tau-Myc-expressing ISNb neurons frequently project in ISNd to muscles 15–17 (arrow). **e, f**, Summary of the wild-type (**e**) and *lim3*-mutant (**f**) phenotypes. In the wild-type embryo, the ISNb (green) and ISNd (red) motor neurons project axons into their correct target areas. In *lim3* mutants, ISNb motor neurons project ectopically in ISNd.

(Fig. 5b). In 21% of hemisegments ($n = 85$) there was a complete loss of ISNd innervation, whereas SNa and other ISN projections appeared to have the normal number and morphology of motor axons. In contrast to misexpression of *lim3*, when we expressed *isl* using the same *GAL4* driver we detected no alterations in ISNd projections ($n = 80$). These results indicate that ISNd motor neurons expressing *isl* only may have been re-routed to ISNb when forced to misexpress *lim3*.

To determine the identity of ectopic ISNb motor neurons, we used the lipophilic tracer DiI to backfill motor neurons from the cleft between muscles 13 and 12. Muscle 12 is normally innervated by two ISNb motor neurons, RP5 and V, both of which are located contralaterally in the VNC¹⁴. Our results confirm these findings and show that, in *ftz_{ng}-GAL4.20/+* control embryos, the RP5 and V motor neurons can reproducibly be backfilled from this muscle cleft (Fig. 5c). In rare cases (6%, $n = 31$) we detected an ipsilateral motor neuron, the identity of which is unknown at present. In *lim3*-misexpressing *ftz_{ng}-GAL4.20/UAS-lim3* embryos we frequently backfilled ipsilateral motor neurons (47%, $n = 32$) in addition to RP5 and V motor neurons (Fig. 5d–f). The position and morphology of these ipsilateral cells are consistent with them being ISNd motor neurons that are ectopically innervating the ISNb target area. Not all ISNd motor neurons in every segment appeared to switch their projection to ISNb in response to *lim3* misexpression; this may occur as a result of variability in expression of *GAL4* from the *ftz_{ng}-GAL4.20* transgene. Alternatively, as described above, there may be competition for the limited number of ISNb muscle targets.

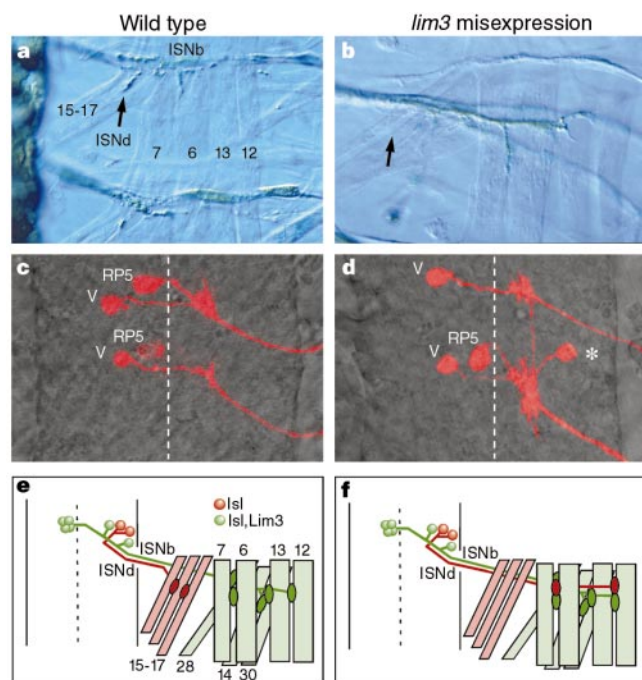


Figure 5 *lim3* misexpression changes motor-neuron projections. **a**, A *ftz_{ng}-GAL4.20/UAS-tau-myc* stage-17 embryo immunostained for Tau-Myc. The *ftz_{ng}-GAL4.20* driver directs expression of *tau-myc* in most, if not all, motor neurons, including those of the ISNd (arrow). **b**, A *ftz_{ng}-GAL4.20/UAS-lim3* stage-17 embryo immunostained for fasciclin II. The ISNd has fewer motor axons projecting in the clefts of muscles 15–17 (arrow) than the wild-type ISNd. Other nerves appear to be unaffected. **c, d**, Embryos in which motor neurons have been backfilled with DiI injected in the cleft between muscles 12 and 13. In *ftz_{ng}-GAL4.20/+* control embryos (**c**) and in wild-type embryos (not shown), the RP5 and V motor neurons are routinely labelled by this procedure¹⁴. **d**, In *lim3*-misexpressing *ftz_{ng}-GAL4.20/UAS-lim3* embryos, further motor neurons, located ipsilaterally (asterisk), are frequently backfilled. **e, f**, Summary of the wild-type (**e**) and the *lim3*-misexpression (**f**) phenotypes. In response to *lim3* misexpression, ISNd motor neurons switch their projections to innervate ISNb muscle targets.

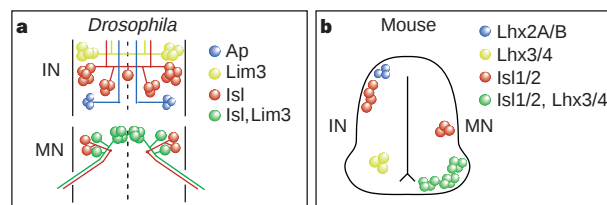


Figure 6 Conservation of expression patterns of LIM-homeodomain proteins between *Drosophila* and mouse. **a**, Expression of *ap*, *isl* and *lim3* in the *Drosophila* VNC compared with **b**, expression of the mouse homologues of these genes, *Lhx2A*, *Lhx2B*, *Lhx3*, *Lhx4*, *Isl1* and *Isl2*, in the spinal cord^{8,22,23}. In both *Drosophila* and mouse, LIM-homeodomain genes are expressed in non-overlapping subsets of interneurons (IN) and are combinatorially expressed by subsets of motor neurons (MN).

These results support the idea that *lim3* and *isl* function as part of a combinatorial code to control the specificity of motor-neuron pathway selection. *Isl* is necessary for the identity of both ISNb and ISNd motor neurons⁴. Our results support a model in which *Lim3*, in combination with *Isl*, subdivides these motor neurons into the individual ISNb and ISNd classes. Mechanistically, this could involve the formation of one or more complexes that are similar to those that form *in vitro* between vertebrate *Isl1*, *Lhx3* and their co-factor NLI/Ldb1 (refs 20, 21). We envision that such complexes might regulate the expression of specific genes involved in repulsion from ISNd targets and/or attraction to ISNb targets. Similarly, other unique combinations of factors are likely to specify other classes of motor neuron. For example, *Isl* and *Lim3* are co-expressed by TN motor neurons in addition to ISNb motor neurons, indicating that an additional factor or factors may be required to differentiate between the TN and ISNb.

The overall expression patterns of *lim3*, *isl* and *ap* in the *Drosophila* VNC are highly analogous to the patterns of expression of their vertebrate counterparts in the developing spinal cord^{1,8,22,23}. In both *Drosophila* and mouse these genes are expressed by non-overlapping subsets of interneurons, and *isl/Isl1/2* and *lim3/Lhx3/4* are expressed combinatorially by motor neurons (Fig. 6). A recent study has shown that mice carrying mutations in both *Lhx3* and *Lhx4*, like *Drosophila* carrying mutations in *lim3*, show an increased number of motor neurons with an *Isl1*-only identity²³. Thus, the conservation of expression of LIM-homeodomain proteins in mice and flies appears to extend to the functional level. □

Methods

Isolation of *lim3* complementary DNA and genomic clones. Degenerate 42-base-pair oligonucleotides, homologous to the third helix of the homeodomain, were used to amplify the following sequences from a *Drosophila* eye-antennal imaginal-disc library: GA(AG)I(CG)(CG)CAG(AG)TCAAGATC TGG, TTCCAGAA(TC)CG(GC)CG(GC)ATGAAG and GA(AG)I(CG)(CG)CAG(AG)TCAAGATCTGGTTCCAGAA(TC)CG(GC)CG(GC)GCCAAG. One of the sequences corresponded to a putative *lim3* homologue, a partial sequence of which had already been reported²⁴. Subsequent screening of an embryonic cDNA library²⁵ yielded a 2.6-kilobase insert, corresponding in size to the *lim3* mRNA according to northern blot analysis. Exon/intron structure was determined by sequence analysis of the 5-kilobase *SalI* fragment derived from EMBL3 genomic clones.

Sequencing of *lim3*-mutant alleles. *lim3* homozygous mutants were identified in embryo collections by the use of the *CyOwg-lacZ* balancer chromosome²⁶. For each mutant allele, we sequenced two different clones, from each of five independent polymerase chain reactions (PCRs), on both strands. The mutations appeared in all five independent PCR reactions; no other areas of the open reading frame showed reproducible sequence alterations. In the *lim3^{Bd1}* cDNA, a T-to-A conversion at position 604 causes a cysteine-to-serine substitution; in *lim3^{Bd7}*, a G-to-A conversion at position 749 causes a cysteine-to-tyrosine substitution; and *lim3^{Bd6}* has a deletion of bases 1,698–1,706.

DiI retrograde labelling. Embryos were dissected and fixed as described²⁷. DiI

(Molecular Probes) was dissolved at 1 mg ml^{-1} in *N,N*-dimethylformamide and ionophoretically injected into the neuromuscular cleft. Embryos were placed in $1 \times \text{PBS}$ overnight at 4°C , mounted in 50% glycerol plus 2% paraformaldehyde, and viewed on a Zeiss confocal microscope.

Generation of transformants. Restriction fragments from *lim3* genomic phage clones⁹ were inserted into P[tau-myc]⁴ or HZ50PL lacZ²⁸ plasmids and transformation was done as described²⁹. Three independent *lim3A-tau-myc* lines and one *lim3A-lacZ* line were tested. All displayed similar expression patterns. Two of the *lim3A-tau-myc* insertions, denoted 1.8 and 1.13, were crossed into *lim3* mutant backgrounds and gave identical results. *UAS-lim3* was made by inserting the *lim3* cDNA plus 50 base pairs of the *Xenopus laevis* β -globin 5' leader from plasmid pNB²⁵ into plasmid pUAS¹⁸. For the mis-expression studies, several independent transgenic lines, each carrying one or two copies of *UAS-lim3*, were used.

The *ftz-ng-GAL4.20* driver. To improve the expression levels of the original *ftz-ng-GAL4* line¹⁹, we generated 25 new insertion lines by transposase-mediated mobilization. Each line was tested for *GAL4* expression by crossing it to flies carrying a *UAS-tau-myc-GFP* reporter³⁰, where *GFP* is green fluorescent protein. One line, *ftz-ng-GAL4.20*, expressed *GAL4* at high levels in most, if not all, motor neurons. As assayed by transactivation of the *UAS-tau-myc* reporter gene, 89% of hemisegments showed labelling in ISNb and 84% showed labelling in ISNd ($n = 56$).

Received 26 October; accepted 8 December 1998.

1. Tsuchida, T. *et al.* Topographic organization of embryonic motor neurons defined by expression of LIM homeobox genes. *Cell* **79**, 957–970 (1994).
2. Appel, B. *et al.* Motoneuron fate specification revealed by patterned LIM homeobox gene expression in embryonic zebrafish. *Development* **121**, 4117–4125 (1995).
3. Way, J. C. & Chalfie, M. *mec-3*, a homeobox-containing gene that specifies differentiation of the touch receptor neurons in *C. elegans*. *Cell* **54**, 5–16 (1988).
4. Thor, S. & Thomas, J. B. The *Drosophila* *islet* gene governs axon pathfinding and neurotransmitter identity. *Neuron* **18**, 397–409 (1997).
5. Lundgren, S. E., Callahan, C. A., Thor, S. & Thomas, J. B. Control of neuronal pathway selection by the *Drosophila* LIM homeodomain gene *apterous*. *Development* **121**, 1769–1773 (1995).
6. Hobert, O., D'Alberti, T., Liu, Y. & Ruvkun, G. Control of neural development and function in a thermoregulatory network by the LIM homeobox gene *lin-11*. *J. Neurosci.* **18**, 2084–2096 (1998).
7. Hobert, O. *et al.* Regulation of interneuron function in the *C. elegans* thermoregulatory pathway by the *trx-3* LIM homeobox gene. *Neuron* **19**, 345–357 (1997).
8. Pfaff, S. L., Mendelsohn, C. L., Stewart, C. L., Edlund, T. & Jessell, T. M. Requirement for LIM homeobox gene *Isl-1* in motor neuron generation reveals a motor neuron-dependent step in interneuron differentiation. *Cell* **84**, 309–320 (1996).
9. Wright, T. R. F. in *Results and Problems in Cell Differentiation* (ed. Hennig, W.) 95–120 (Springer, Heidelberg, 1987).
10. Lindley, D. L. & Zimm, G. G. *The Genome of Drosophila melanogaster* (Academic, San Diego, 1992).
11. Michelsen, J. W., Schmeichel, K. L., Beckerle, M. C. & Winge, D. R. The LIM motif defines a specific zinc-binding protein domain. *Proc. Natl Acad. Sci. USA* **90**, 4404–4408 (1993).
12. Michelsen, J. W. *et al.* Mutational analysis of the metal sites in an LIM domain. *J. Biol. Chem.* **269**, 11108–11113 (1994).
13. Surdej, P. & Jacobs-Lorena, M. Developmental regulation of *bicoid* mRNA stability is mediated by the first 43 nucleotides of the 3' untranslated region. *Mol. Cell. Biol.* **18**, 2892–2900 (1998).
14. Landgraf, M., Bossing, T., Technau, G. M. & Bate, M. The origin, location, and projections of the embryonic abdominal motoneurons of *Drosophila*. *J. Neurosci.* **17**, 9642–9655 (1997).
15. Sink, H. & Whittington, P. M. Location and connectivity of abdominal motoneurons in the embryo and larva of *Drosophila melanogaster*. *Neurobiol.* **22**, 298–311 (1991).
16. Jan, L. Y. & Jan, Y. N. Antibodies to horseradish peroxidase as specific neuronal markers in *Drosophila* and grasshopper embryos. *Proc. Natl Acad. Sci. USA* **79**, 2700–2704 (1982).
17. Van Vactor, D., Sink, H., Fambrough, D., Tsou, R. & Goodman, C. S. Genes that control neuromuscular specificity in *Drosophila*. *Cell* **73**, 1155–1164 (1993).
18. Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).
19. Lin, D. M. & Goodman, C. S. Ectopic and increased expression of Fasciclin II alters motoneuron growth cone guidance. *Neuron* **13**, 507–523 (1994).
20. Jurata, L. W., Pfaff, S. L. & Gill, G. N. The nuclear LIM domain interactor NLI mediates homo- and heterodimerization of LIM domain transcription factors. *J. Biol. Chem.* **273**, 3152–3157 (1998).
21. Morcillo, P., Rosen, C., Baylies, M. K. & Dorsett, D. Chip, a widely expressed chromosomal protein required for segmentation and activity of a remote wing margin enhancer in *Drosophila*. *Genes Dev.* **11**, 2729–2740 (1997).
22. Liem, K. F. J., Tremml, G. & Jessell, T. M. A role for the roof plate and its resident TGF β -related proteins in neuronal patterning in the dorsal spinal cord. *Cell* **91**, 127–138 (1997).
23. Sharma, K. *et al.* Lhx3 and Lhx4 assign identities for motor neurons projecting axons ventral from the neural tube. *Cell* (in the press).
24. Kalionis, B. & O'Farrell, P. H. A universal target sequence is bound *in vitro* by diverse homeodomains. *Mech. Dev.* **43**, 57–70 (1993).
25. Brown, N. H. & Kafatos, F. C. Functional cDNA libraries from *Drosophila* embryos. *J. Mol. Biol.* **203**, 425–437 (1988).
26. O'Keefe, D. D., Thor, S. & Thomas, J. B. Function and specificity of LIM domains in *Drosophila* nervous system and wing development. *Development* **125**, 3915–3923 (1998).
27. Callahan, C. A. & Thomas, J. B. Tau- β -galactosidase, an axon-targeted fusion protein. *Proc. Natl Acad. Sci. USA* **91**, 5972–5976 (1994).
28. Hiromi, Y. & Gehring, W. J. Regulation and function of the *Drosophila* segmentation gene *fushi tarazu*. *Cell* **50**, 963–974 (1987).
29. Rubin, G. M. & Spradling, A. C. Genetic transformation of *Drosophila* with transposable element vectors. *Science* **218**, 348–353 (1982).
30. Callahan, C. A., Yoshikawa, S. & Thomas, J. B. Tracing axons. *Curr. Biol.* **8**, 582–586 (1998).

Acknowledgements. We thank C. S. Goodman for the *ftz-ng-GAL4* flies and anti-fasciclin II antibody; S. Yoshikawa for the *UAS-tau-myc-GFP* flies; the Midamerica, Umea and Bloomington Stock Centers for providing fly stocks; L. W. Jurata and G. N. Gill for sharing unpublished results; K. Bishop and members of our laboratory for providing advice and technical assistance; and S. Pfaff, K. Sharma, G. Lemke and D. van Meyel for critical review of the manuscript. This work was supported by grants from the NIH and a Pew Scholars Award from the Pew Memorial Trusts (to J.B.T.), an EMBO Long Term Fellowship (to S.G.E.A.) and an HFSP Long-term Fellowship (to S.T.).

Correspondence and requests for materials should be addressed to S.T. The *lim3* gene sequence has been deposited with GenBank under accession number AF 109306.

A protochlorophyllide light-harvesting complex involved in de-etiolation of higher plants

Christiane Reinbothe*, Nikolai Lebedev† & Steffen Reinbothe*

* Université Joseph Fourier et Centre National de la Recherche Scientifique, BP53, CERMO, F-38041 Grenoble cedex 9, France

† Department of Biology, Gilmer Hall, University of Virginia, Charlottesville, Virginia 22903-2477, USA

When etiolated angiosperm seedlings break through the soil after germination, they are immediately exposed to sunlight, but at this stage they are unable to perform photosynthesis¹. In the absence of chlorophyll *a* and chlorophyll *b*, two other porphyrin species cooperate as the basic light-harvesting structure of etiolated plants. Protochlorophyllide *a* and protochlorophyllide *b* (ref. 2) form supramolecular complexes with NADPH and two closely related NADPH:protochlorophyllide oxidoreductase (POR) proteins—PORA and PORB (ref. 3)—in the prolamellar body of etioplasts. Here we report that these light-harvesting POR–protochlorophyllide complexes, named LHPP, are essential for the establishment of the photosynthetic apparatus and also confer photoprotection on the plant. They collect sunlight for rapid chlorophyll *a* biosynthesis and, simultaneously, dissipate excess light energy in the bulk of non-photo-reducible protochlorophyllide *b*. Based on this dual function, it seems that LHPP provides the link between skotomorphogenesis and photosynthesis that is required for efficient de-etiolation.

Previous studies have shown that the prolamellar body, which is present in etioplasts of dark-grown angiosperm seedlings, contains several different spectral forms of protochlorophyllide (Pchl)ide⁴. The main Pchl species with an absorption maximum at 650 nm and an emission maximum at 657 nm, Pchl_{650–657}, was capable of being converted into chlorophyllide (Chl)ide_{684–690} when illuminated with a single white-light flash⁴.

We re-examined these spectral forms of Pchl and Chl in the prolamellar body of etiolated barley seedlings before and after a 1-ms white-light flash. Fluorescence emission spectroscopy recorded at low temperature (77K) revealed that both Pchl_{650–657} and Chl_{684–690} could be excited at 470 nm (Fig. 1a). Excitation at 440 nm, however, led to two fluorescence peaks of Pchl, one at 633 nm and another at 657 nm, of which only the Pchl₆₅₇ was converted to Chl₆₉₀ when flash-light-illuminated (Fig. 1b). Consistent with previous findings⁴, this observation indicated that part of the Pchl might be present in a free, non-photo-reducible form (Pchl₆₃₃). Room-temperature fluorescence emission and excitation analyses of acetone-extracted pigments (Fig. 1c, d, respectively) supported this hypothesis and demonstrated the occurrence of only one major species of Pchl. However, this form of Pchl displayed features that were reminiscent of both Pchl *a* and Pchl *b* (ref. 2) (Fig. 1c, d; see ref. 5 for details on their biochemistry). Furthermore, electrophoretic studies showed

that there were two POR-related polypeptides, previously identified as PORA and PORB (ref. 3), in membrane preparations of the prolamellar body of barley etioplasts (Fig. 1e). These PORA and PORB proteins were expressed differentially upon illumination of etiolated seedlings (Fig. 1e).

Based on the contradictory results provided by spectroscopic and electrophoretic analyses, we attempted to determine the actual functions of PORA and PORB in the prolamellar body of barley etioplasts. We assumed that PORA and PORB may form oligomeric complexes designed for light-harvesting and subsequent chlorophyll (Chl) *a* biosynthesis that is necessary when the etiolated plant emerges from the soil. By analogy to LHCII, the principal light-harvesting complex of higher-plant photosynthesis^{6–8}, light absorbed by the chlorophyll *b* precursor Pchl *b* could be transferred to Pchl *a*. In LHCII, energy transfer from Chl *b* to the closely related Chl *a* (both compounds differ only in a formyl group replacing a methyl group at the 7-position in the chlorin ring of Chl *a*; ref. 5) can take place because of the different energy contents and decay times of their excited states^{7,8}. Energy absorbed by Chl *b* is transferred to Chl *a* within less than 1 ps, where it remains for 1–3 ns, as shown by the Chl *a* fluorescence decay time in the purified, detergent-solubilized complex^{7,8}. We assumed that a similar mechanism could operate in the case of Pchl *b* and Pchl *a*. As a result, Pchl *a* would be reduced to Chlide *a*.

To demonstrate that a putative Pchl *b*-based light-harvesting structure was operating, two prerequisites needed to be met. First, PORA and PORB should possess different substrate specificities

for Pchl *a* and Pchl *b*. Second, PORA–pigment and PORB–pigment complexes should be able to form supramolecular structures in which Pchl *b* could accomplish the anticipated light-harvesting antenna function.

To test the first assumption, PORA and PORB polypeptides were generated by an approach based on the polymerase chain reaction (PCR)⁹, synthesized by coupled *in vitro* transcription/translation¹⁰ of the respective recombinant clones, purified¹¹, and reconstituted to PORA–pigment–NADPH and PORB–pigment–NADPH complexes¹². Instead of using Pchl *a* and Pchl *b*, we used their zinc (Zn) analogues, Zn protoporphorbide (ZnPP) *a* (ref. 13) and ZnPP *b* (ref. 13), respectively (Pchl *a* and *b* contain magnesium (Mg) instead of Zn; see ref. 13 for the structural formulas). The main reason for using Zn analogues was that these compounds can be synthesized in sufficient amounts and in a chemically pure form¹³. Protochlorophyllide *a* and Pchl *b*, however, are always present in variable proportions in etiolated barley tissues (C.R. and S.R., unpublished results) and are less stable than their Zn derivatives¹³. Furthermore, POR protein preparations of etiolated wheat plants had previously been shown to accept ZnPP *a* and ZnPP *b* as substrates¹³.

When the levels of POR-bound pigments were analysed spectrometrically after extraction with acetone¹⁴, striking differences were observed: PORA bound approximately tenfold more ZnPP *b* than ZnPP *a* (Fig. 2a, solid versus dashed lines). Conversely, PORB preferred ZnPP *a* and bound approximately tenfold less ZnPP *b* (Fig. 2b, dashed versus solid lines).

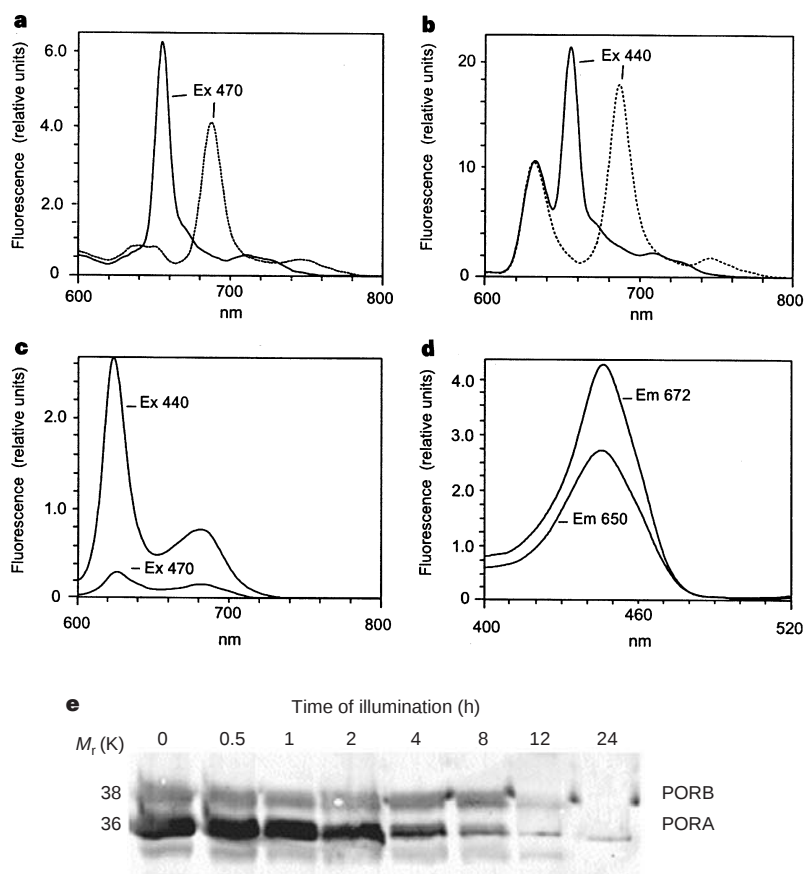


Figure 1 Authentic LHPP contains two POR proteins but only one photoactive species of Pchl *b*. **a**, Low-temperature (77 K) fluorescence emission spectra at an excitation wavelength of 470 nm (Ex 470) of porphyrins present in prolamellar-body membranes of etioplasts before (solid line) and after (dashed line) a single, 1-ms white-light flash. **b**, As **a**, but recorded at 440 nm (Ex 440). **c**, Room-temperature fluorescence emission analysis at 440 nm (Ex 440) and 470 nm (Ex

470) of acetone-extracted, non-flashed prolamellar-body membranes. **d**, Fluorescence excitation spectra at emission wavelengths of 650 nm (Em 650) and 672 nm (Em 672), respectively, of the same sample as in **c**. **e**, Western blot analysis²⁴ of POR-related proteins²¹ in isolated prolamellar body membranes (0 h) and membrane preparations from white-light-illuminated (in hours) plants.

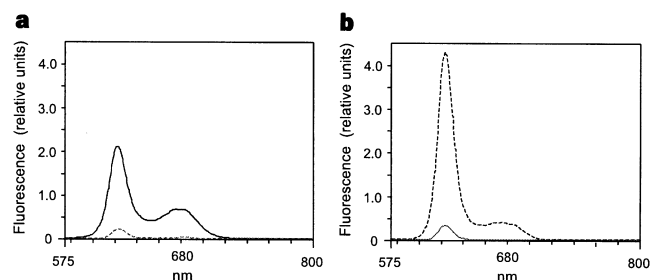


Figure 2 PORA and PORB bind ZnPPa and ZnPPb, the Zn analogues of Pchlide *a* and Pchlide *b*, with different affinities. **a**, Room-temperature fluorescence emission analysis at 440 nm of acetone-extracted PORA-ZnPPb-NADPH (solid line) and PORA-ZnPPa-NADPH (dashed line) complexes. **b**, As **a**, but with PORB-ZnPPb-NADPH (solid line) and PORB-ZnPPa-NADPH (dashed line) complexes. ZnPPa has a maximum fluorescence emission at 631 nm, whereas ZnPPb emits at 627 nm (ref. 13).

To test the second assumption, PORA-ZnPPb-NADPH and PORB-ZnPPa-NADPH complexes that had been reconstituted *in vitro* as described were incubated with each other in the dark for 15 min (Fig. 3a). The incubation mixture was then subjected to gel filtration on Sephadex G100. When oligomer formation was assayed electrophoretically, two fractions were positive (fractions 5 and 15 in Fig. 3b). In either case, a 5:1 stoichiometry of PORA-ZnPPb-NADPH to PORB-ZnPPa-NADPH was observed. After extracting the gel-filtered, oligomeric complexes contained in fraction 5 with acetone, room-temperature fluorescence spectroscopy performed at an excitation wavelength of 440 nm demonstrated a major fluorescence peak at 631 nm (Fig. 4a). When these oligomeric complexes were illuminated before acetone extraction, their spectral properties changed. The fluorescence peak at 631 nm decreased (Fig. 4b).

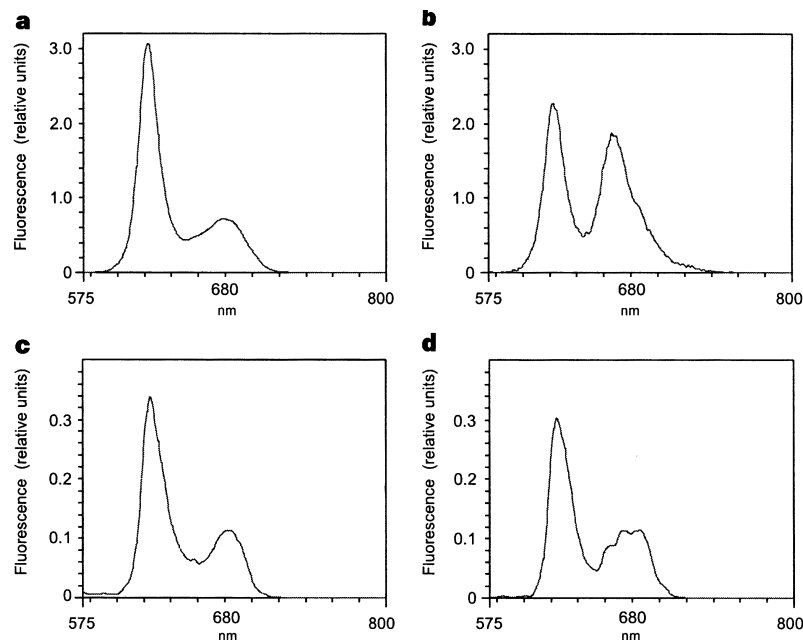


Figure 4 Reconstituted LHPP catalyses the reduction of ZnPPa to Zn pheophorbide *a*. **a-d**, LHPP was reconstituted *in vitro* and purified by gel filtration, as described in Fig. 3. Reconstituted LHPP recovered from fraction 5 was then either kept in the dark (**a, c**) or exposed to white light (**b, d**) for 15 min. Pigments were subsequently extracted with acetone and their fluorescence

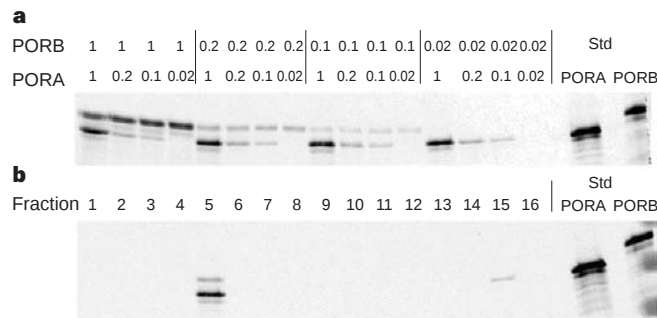


Figure 3 Reconstitution of LHPP *in vitro*. **a**, Autoradiogram showing purified, electrophoretically resolved PORA-ZnPPb-NADPH and PORB-ZnPPa-NADPH complexes corresponding to 1, 0.2, 0.1 and 0.02 enzyme equivalents as indicated. As standards (Std), gel-filtered PORA-ZnPPb-NADPH and PORB-ZnPPa-NADPH were analysed individually. **b**, Detection of oligomer formation between PORA-ZnPPb-NADPH and PORB-ZnPPa-NADPH, that is, reconstituted LHPP, by gel filtration on Sephadex G100 and subsequent SDS-polyacrylamide gel electrophoresis.

Simultaneously, a new fluorescence peak appeared at 665 nm (Fig. 4b), suggesting that ZnPPb bound to PORA had absorbed light and transferred the excitation energy onto the PORB-ZnPPa-NADPH core complex, where the reduction of ZnPPa to Zn pheophorbide *a* ultimately occurred.

Indeed, re-recording the emission spectra at an excitation wavelength of 470 nm (Fig. 4c, d), which has previously been shown to excite exclusively ZnPPb and Zn pheophorbide *b*, but not ZnPPa and its respective product¹³, confirmed that only PORB had been catalytically active in the reconstituted POR-pigment complex. We tentatively named this complex LHPP—light-harvesting POR-Pchlide complex. There was no fluorescence emission at 652 nm, which is indicative of the operation of PORA reducing ZnPPb to Zn pheophorbide *b* (Fig. 4c, d) in LHPP.

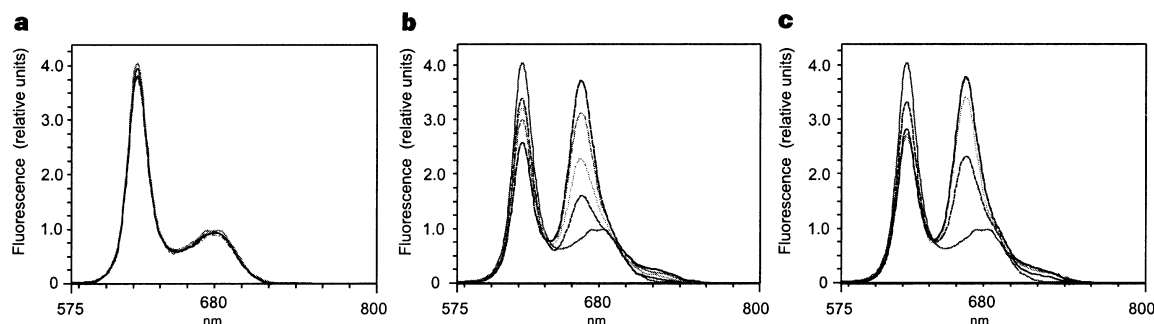


Figure 5 Reconstituted LHPP is a light-harvesting complex, collecting solar energy for the formation of PORB-derived Zn pheophorbide *a*. **a–c**, LHPP was reconstituted *in vitro* and purified by gel filtration as described in Fig. 3b (fraction 5), and the sample divided into three equal parts. One part was kept in the dark (**a**), another was exposed to a narrow 470-nm light beam (**b**), and the third part was

exposed to continuous white light (**c**). After 0 (solid line), 2 (short-dashed line), 5 (dotted line), 10 (dot-dashed line) and 15 (long-dashed line) min of incubation, aliquots of each of the different samples were extracted with acetone and their fluorescence emission properties determined at room temperature at 440 nm.

To demonstrate that PORA had indeed been inactive but served a light-harvesting function, LHPP was reconstituted *in vitro* in the dark and illuminated with a narrow 470-nm light beam. This light treatment caused the conversion of ZnPP*a* to Zn pheophorbide *a*, as seen by the decrease in fluorescence emission at 631 nm and the increase in fluorescence emission at 665 nm in samples extracted with acetone (Fig. 5b). Compared with reconstituted LHPP illuminated with white light (Fig. 5c), product formation was delayed, presumably because of the lower quantum flux provided by the narrow 470-nm light beam. In dark controls, no Zn pheophorbide *a* formation was observed (Fig. 5a).

Compared with authentic LHPP present in membranes from the prolamellar body of etioplasts, the physiological properties of reconstituted LHPP appeared to be identical. But no Pchl*ide*_{650–657} could be traced when the fluorescence properties of reconstituted LHPP were analysed at low temperature (Fig. 6a, solid line). We assumed that this lack of Pchl*ide*_{650–657} might be because reconstituted LHPP, in contrast to authentic LHPP, did not contain lipids that may interact with PORA and PORB to influence the spectroscopic properties of their bound pigments¹⁵. To test this we purified a mixed galactolipid and sulpholipid fraction (monogalactosyl diacylglycerol:digalactosyl diacylglycerol:sulphoquinovosyl diacylglycerol, 58:36:6 mol%) from prolamellar bodies of barley etioplasts¹⁶ and incubated this fraction with reconstituted LHPP. The fluorescence properties of LHPP were then examined by low-temperature spectroscopy at 440 nm, as described previously.

Addition of the mixed galacto- and sulpholipids indeed gave rise

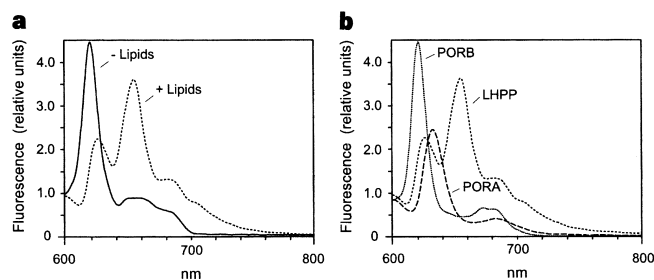


Figure 6 Mixed galacto- and sulpholipids promote the formation of Pchl*ide*_{650–657} in reconstituted LHPP. **a**, LHPP, reconstituted as described in Fig. 3b (fraction 5), was incubated with a mixed galactolipid and sulpholipid preparation from barley etioplasts (dashed line, + lipids). Control incubations were identical except for lacking lipids (solid line, – lipids). **b**, As **a**, but with PORA–ZnPP*b*–NADPH complexes (long dashes) and PORB–ZnPP*a*–NADPH complexes (dotted line) incubated individually with the mixed galacto- and sulpholipid preparation. Note that neither POR–pigment complex is able to give rise to Pchl*ide*_{650–657}, in contrast to reconstituted LHPP shown for comparison (dashed line).

to Pchl*ide*_{650–657} (Fig. 6a, dotted line). By contrast, neither PORA–ZnPP*b*–NADPH nor PORB–ZnPP*a*–NADPH complexes, when tested individually, were able to form Pchl*ide*_{650–657} (Fig. 6b).

Functional LHPP consists of five PORA–Pchl*ide b*–NADPH and one PORB–Pchl*ide a*–NADPH ternary complexes, presumably held together in supramolecular, ring-like structures, as deduced from electron micrographs of prolamellar-body membrane sections¹⁵. An analogous structure has been reconstituted *in vitro*, using PORA and PORB polypeptides synthesized from corresponding complementary DNA clones, the Zn analogues of Pchl*ide b* and Pchl*ide a*, plus NADPH. Energy transfer in the authentic Mg-containing LHPP and in the reconstituted, Zn-containing LHPP, was virtually identical and proceeded from Pchl*ide b* bound to PORA to Pchl*ide a*. This was reduced to Chl*ide a*, by virtue of PORB in the core complex. In both the prolamellar body of etioplasts and in the reconstituted system, the spectral properties of LHPP depended critically on the presence of galacto- and sulpholipids. Owing to the presence of these lipids, Pchl*ide*_{650–657} and its Zn-containing counterpart became the main photoconvertible species of Pchl*ide*. Protochlorophyllide (ZnPP) *b* bound to PORA in the periphery of either authentic LHPP or reconstituted LHPP was not photoconvertible in the first instance. Rather, it absorbed light and transferred the excitation energy onto PORB. Excess light energy was emitted as fluorescence, suggesting an additional association of LHPP with carotenoids in the prolamellar body of etioplasts. Further to its function as a light scavenger, PORA seems to operate as a Pchl*ide*-reducing enzyme. When PORA is released from the prolamellar body, it converts Pchl*ide b* to Chl*ide b*. This reaction operates for only a short period after de-etiolation, after which the expression of PORA is switched off^{3,17} (Fig. 1e). PORB, by contrast, is constitutively expressed (Fig. 1e) and drives Chl*ide a* synthesis not only in etiolated plants at the beginning of illumination but also in green plants^{3,17}. In either case, the PORB-derived Chl*ide a* in turn is esterified with phytol to Chl *a*, which initiates chloroplast development by controlling the intraplasmic synthesis and assembly of the photosynthetic complexes in the thylakoid membranes, including the plastid-encoded Chl-*a*-binding proteins of the reaction centres and the nuclear-encoded, post-translationally imported LHC proteins^{5,18}. LHPP-driven Chl*ide a* synthesis is hence a key process in the greening of etiolated angiosperm plants and a switch from skotomorphogenesis to photosynthetic growth. In this context, it is remarkable to note that the action spectra of phytochrome-controlled hypocotyl elongation and LHPP-driven Chl*ide a* synthesis overlap at 650 nm (ref. 19). This is consistent with our results because skotomorphogenesis is regarded as a developmental strategy through which the plant prepares for light-harvesting when emerging from the soil under conditions of light limitation²⁰. Work is in progress to further characterize LHPP. □

Methods

In situ and in vitro fluorescence measurements of pigments. Etioplasts were prepared from dark-grown seedlings of barley (*Hordeum vulgare* cv. Carina) as described¹². Similarly, plastids were isolated from etiolated seedlings that had additionally been exposed to light for various time intervals. Prolamellar bodies recovered from isolated etioplasts were subsequently subjected to low-temperature fluorescence spectroscopy at excitation wavelengths of either 470 nm or 440 nm in a spectrometer LS50 (Perkin Elmer Corp., Norwalk, CT). After a single 1-ms white-light flash, the samples were re-recorded under the same conditions. Prolamellar-body membranes from parallel assays were extracted with acetone. Fluorescence emission spectroscopy was done at room temperature at excitation wavelengths of either 470 nm or 440 nm, whereas fluorescence excitation spectroscopy was done at 23 °C at emission wavelengths of either 650 nm or 672 nm.

Reconstitution of POR-pigment complexes and LHPP. Double-stranded DNAs encoding the mature PORA and PORB of barley were produced by a PCR-based approach⁹, using primers 1 (5'-AACTGCAGATGGGCAAGAAGACGCTGCGGCAG3') plus 2 (5'-AACTGCAGGTTGGATCATAGTCCGACGAGCTT3'), and primers 3 (5'-AACTGCAGATGGGCAAGAAGACTGTCCGACG3') plus 4 (5'-AACTGCAGTATCATGCGAGCCGACGAGCTT3'), and cDNA clones A7 (ref. 21) and L2 (ref. 3), respectively, as templates. Radiolabelled PORA and PORB molecules synthesized by coupled *in vitro* transcription/translation¹⁰ of respective recombinant clones were subsequently reconstituted to enzyme-pigment complexes¹². Protein-bound ZnPPa and ZnPPb recovered after gel filtration on Sephadex G15 were extracted with acetone and detected at room temperature by their fluorescence emission maxima at 631 nm and 627 nm, respectively, at an excitation wavelength of 440 nm¹³. For reconstitution of LHPP, PORA-ZnPPb-NADPH and PORB-ZnPPa-NADPH ternary complexes corresponding to 1, 0.2, 0.1 and 0.02 enzyme equivalents were incubated with each other in the dark for 15 min. Half each of the incubation mixtures was immediately precipitated with trichloroacetic acid, whereas the other halves were subjected to gel filtration on Sephadex G100. Proteins running in the flowthrough were precipitated with trichloroacetic acid. After their further processing with ethanol and diethyl ether, the different proteins were resolved on 11–20% polyacrylamide gradients containing SDS and detected by autoradiography²².

Preparation of lipids. Prolamellar-body membranes were isolated from dark-grown barley seedlings as described¹² and illuminated with a series of 15 successive 1-ms light flashes. After the disintegration of the prolamellar body had been completed, as assessed microscopically, the resuspended membranes were incubated for an additional 30-min period to induce PORA-driven Chlide *b* formation. A POR-degrading protease activity, isolated from barley chloroplasts²³, was then added. Pigment-free, mixed galacto- and sulpholipids were prepared as described¹⁷ and added to LHPP that had been reconstituted *in vitro* as specified above.

Received 14 April; accepted 3 November 1998.

- Granick, S. The structural and functional relationship between heme and chlorophyll. *Harvey Lect.* **44**, 220–245 (1950).
- Shedibalkar, V. P., Ioannides, I. M. & Rebeiz, C. A. Chloroplast biogenesis. Detection of monovinyl protochlorophyllide *b* in plants. *J. Biol. Chem.* **266**, 17151–17157 (1991).
- Holtorf, H., Reinbothe, S., Reinbothe, C., Berez, B. & Apel, K. Two routes of chlorophyllide synthesis that are differentially regulated by light in barley. *Proc. Natl Acad. Sci. USA* **92**, 3254–3258 (1995).
- Griffiths, W. T. Reconstitution of chlorophyll formation by isolated etioplast membranes. *Biochem. J.* **174**, 681–692 (1978).
- Reinbothe, S. & Reinbothe, C. The regulation of enzymes involved in chlorophyll biosynthesis. *Eur. J. Biochem.* **237**, 323–343 (1995).
- Kühlbrandt, W., Wang, D. N. & Fujiyoshi, Y. Atomic model of plant light-harvesting complex by electron crystallography. *Nature* **367**, 614–621 (1994).
- Palsson, L. O., Spangfort, M. D., Gulbinas, V. & Gillbro, T. Ultrafast chlorophyll *b* to chlorophyll *a* excitation energy transfer in the isolated light harvesting complex, LHCl2, of green plants: implications for the organisation of chlorophylls. *FEBS Lett.* **339**, 134–138 (1994).
- Ide, J. P., Klug, D. R., Kühlbrandt, W., Georgi, L. & Porter, G. The state of detergent-solubilized light-harvesting chlorophyll-*a/b* protein complex as monitored by picosecond time-resolved fluorescence and circular dichroism. *Biochim. Biophys. Acta* **893**, 349–364 (1987).
- Innis, M. A., Gelfand, D. H., Sninsky, J. J. & White, T. J. *PCR Protocols* (Academic, San Diego, 1990).
- Krieg, P. A. & Melton, D. A. Functional messenger RNAs are produced by SP6 *in vitro* transcription of cloned DNAs. *Nucleic Acids Res.* **12**, 7057–7070 (1984).
- Apel, K., Santel, H.-J., Redlinger, T. E. & Falk, H. The protochlorophyllide holochrome of barley. Isolation and characterization of the NADPH:protochlorophyllide oxidoreductase. *Eur. J. Biochem.* **111**, 251–258 (1980).
- Reinbothe, S., Runge, S., Reinbothe, C., van Cleve, B. & Apel, K. Substrate-dependent transport of the NADPH:protochlorophyllide oxidoreductase into isolated plastids. *Plant Cell* **7**, 161–172 (1995).

- Schoch, S., Helfrich, M., Wiktorsson, B., Sundqvist, C., Rüdiger, W. & Ryberg, M. Photoreduction of protoporphyrin with NADPH-protochlorophyllide oxidoreductase from etiolated wheat (*Triticum aestivum*). *Eur. J. Biochem.* **229**, 291–298 (1995).
- Lebedev, N., van Cleve, B., Armstrong, G. A. & Apel, K. Chlorophyll synthesis in a de-etiolated (det340) mutant of *Arabidopsis* without NADPH-protochlorophyllide (Pchl)ide oxidoreductase (POR) A and photoactive Pchlide-P655. *Plant Cell* **7**, 2081–2090 (1995).
- Ryberg, M. & Sundqvist, C. In *Chlorophylls* (ed. Scheer, H.) 587–612 (CRC Press, Boca Raton, 1991).
- Ryberg, M., Sandelius, A. S. & Selstam, E. Lipid composition of prolamellar bodies and prothylakoids of wheat etioplasts. *Physiol. Planta* **57**, 555–560 (1983).
- Armstrong, G. A., Runge, S., Frick, G., Sperling, U. & Apel, K. Identification of protochlorophyllide oxidoreductases A and B: a branched pathway for light-dependent chlorophyll biosynthesis in *Arabidopsis thaliana*. *Plant Physiol.* **108**, 1505–1517 (1995).
- von Wettstein, D., Gough, S. & Kannangara, C. G. Chlorophyll biosynthesis. *Plant Cell* **7**, 1039–1057 (1995).
- Quail, P. H., Boylan, M. T., Parks, B. M., Short, T. W., Xu, Y. & Wagner, D. Phytochromes: photosensory perception and signal transduction. *Science* **268**, 675–680 (1995).
- Kendrick, R. E. & Kronenberg, H. M. (eds) *Photomorphogenesis in Plants* (Kluwer, Dordrecht, 1991).
- Schulz, R. *et al.* Nucleotide sequence of a cDNA coding for the NADPH-protochlorophyllide oxidoreductase (POR) of barley (*Hordeum vulgare* L.) and expression in *Escherichia coli*. *Mol. Gen. Genet.* **217**, 355–361 (1989).
- Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685 (1970).
- Reinbothe, C., Apel, K. & Reinbothe, S. A light-induced protease from barley plastids degrades NADPH:protochlorophyllide oxidoreductase complexed with chlorophyllide. *Mol. Cell. Biol.* **15**, 6206–6212 (1995).
- Towbin, M., Staehelin, T. & Gordon, J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl Acad. Sci. USA* **76**, 4350–4354 (1979).

Acknowledgements. We thank S. Schoch for a gift of ZnPPa and ZnPPb. This work is dedicated to R. Mache on the occasion of his 65th birthday.

Correspondence and requests for materials should be addressed to C.R. (e-mail: christiane.reinbothe@uni-bayreuth.de).

The imprinting box of the mouse *Igf2r* gene

Yehudit Birger, Ruth Shemer, Jonathan Perk & Aharon Razin

Department of Cellular Biochemistry, The Hebrew University Hadassah Medical School, PO Box 12272, Jerusalem 91120, Israel

Genomic imprinting is a phenomenon characterized by parent-of-origin-specific expression. The imprint is a mark established during germ-cell development to distinguish between the paternal and maternal copies of the imprinted genes. This imprint is maintained throughout embryo development and erased in the embryonic gonads to set the stage for a new imprint¹. DNA methylation is essential in this process as shown by the presence of differentially methylated regions (DMRs) in all imprinted genes² and by the loss of imprinting in mice that are deficient in DNA methylation³ or upon deletion of DMRs^{4–6}. Here we show that a DMR in the imprinted *Igf2r* gene (which encodes the receptor for insulin-like growth factor type-2) that has been shown to be necessary for imprinting⁵ includes a 113-base-pair sequence that constitutes a methylation imprinting box. We identify two new *cis*-acting elements in this box that bind specific proteins: a *de novo* methylation signal and an allele-discrimination signal. We propose that this regulatory system, which we show to be involved in the establishment of differential methylation in the *Igf2r* DMR, represents a critical element in the imprinting process.

The mouse *Igf2r* gene, which is located on chromosome 17 (ref. 7), is maternally expressed, starting at 6.5 days postcoitum⁸. This gene has two DMRs: DMR1, which is located at the promoter region, and DMR2, an intronic 3-kilobase (kb) CpG island (a region of DNA with a high G + C content) (Fig. 1a). DMR2 is maternally methylated starting at the final stages of egg maturation^{7,9,10}.

Our initial studies focused on a single methylation site in this region, *HpaII* site 4 (H4). This site is unmethylated in both oocyte and sperm, and undergoes prompt methylation in the maternal pronucleus of the fertilized egg before syngamy. This methylation is maintained throughout development of the preimplantation embryo¹⁰. To define the minimal sequence required for establishing

the monoallelic methylation of site H4, we used our previously developed method whereby DNA fragments are injected into the male or female pronucleus followed by polymerase chain reaction (PCR) analysis of the methylation status of specific CpG sites in the originally injected and integrated DNA¹¹ (see Methods for details).

When the entire DMR2 was injected into the female pronucleus, site H4 underwent *de novo* methylation which was completed by the eight-cell stage. The same fragment injected into the male pronucleus remained unmethylated (Fig. 1b). This observation is consistent with maternal methylation of this site in the endogenous gene^{7,10}. In control experiments, fragments of the non-imprinted genes *APOAI* (encoding apolipoprotein AI) and *PGK1* (encoding phosphoglycerate kinase-1) were injected into the male or female pronucleus. *HpaII* sites in these DNA sequences remained unmethylated, regardless of the pronucleus into which these fragments were injected (Fig. 1c). The observed parent-specific *de novo* methylation of H4 implies that the injected sequence is recognized and marked as a sequence that should be differentially methylated. Our results indicate that the capacity to introduce the initial mark on the sequence is present in the pronucleus and that the integration and marking of the injected fragment takes place before syngamy. The somewhat late *de novo* methylation (between the four- and eight-cell stage) implies that the mark persists on the DNA for at least two or three cell divisions until methylation is established. It is possible that the differential recognition of this sequence in the male and female pronucleus is attributed to differences in chromatin structure. The different chromatin structure in the egg and sperm is gradually erased during early embryo development, but some of the differences may still exist when differential methylation is first observed.

We used deletion analysis to identify the *cis* elements that control this monoallelic *de novo* methylation. The minimal sequence that could support allele-specific *de novo* methylation is the 113-base-pair (bp) fragment described in Fig. 2a. Its injection into the

maternal pronucleus resulted in *de novo* methylation of site H4, whereas injection of the same fragment into the paternal pronucleus did not result in methylation of the site (Fig. 2b, experiment 1). Single-base changes at the 3' end of the 113-bp fragment abolished *de novo* methylation of site H4 (Fig. 2b, experiments 2–4). These experiments and experiment 5 in Fig. 2b enabled us to define the minimal sequence that is sufficient to direct *de novo* methylation of site H4. The 8-bp sequence (TCACCTCG) at the 3' end (Fig. 2a) seems to be the *cis* element that is involved in the recognition of DMR2 as a target for *de novo* methylation, and this element therefore constitutes a *de novo* methylation signal (DNS).

We next attempted to identify the signal responsible for the discrimination between the two alleles that facilitates monoallelic methylation. Single-base substitutions at the 5' end of the 113-bp sequence abolished allele discrimination (Fig. 2b, experiments 6–8). However, mutation of the C residue at position 7 (see Fig. 2a) did not alter the allele-specific effect of the 113-bp sequence (Fig. 2b, experiment 9). These results imply that the 6 bp at the 5' end of the 113-bp sequence (GTAGCC) constitute an allele-discriminating signal (ADS) that prevents the paternal allele from undergoing methylation *de novo*.

Taken together, these results define a 113-bp box that includes an 8-bp DNS at the 3' end and a 6-bp ADS at the 5' end. This conclusion is based on the methylation analysis of only one site

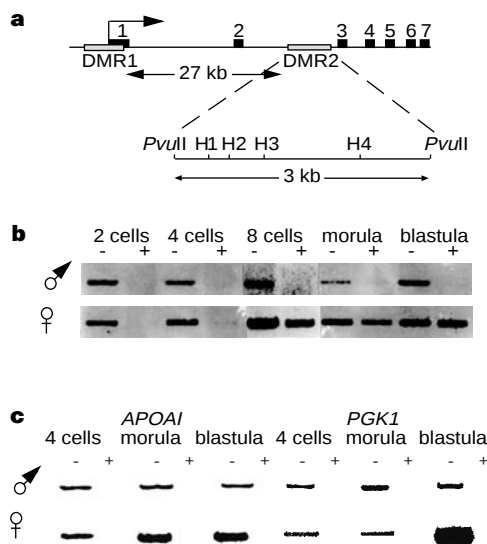


Figure 1 Allele-specific methylation of the *Igf2r* DMR2. **a**, The mouse *Igf2r* gene. Black boxes designate the exons; open boxes designate the two differentially methylated regions, DMR1 and DMR2. *HpaII* sites (H1–H4) in DMR2 have been studied previously¹⁰. **b**, The status of methylation of site H4 in the injected fragment, at various stages of development, after injection of the 3-kb *PvuII* fragment into the male or female pronucleus of the fertilized egg. Genomic DNA samples were digested with (+) or without (–) *HpaII* before amplification of the DNA by PCR (primers 1 and 2) (see Methods). **c**, The injected fragment was: left panel, the 5' 1.8-kb *HindIII*/*PvuII* fragment of the human *APOAI* gene (primers 3 and 4)¹³; right panel, the 0.8-kb *BamHI*/*EcoRI* fragment including exon 1 of the human *PGK1* gene¹⁶ (primers 5 and 6).

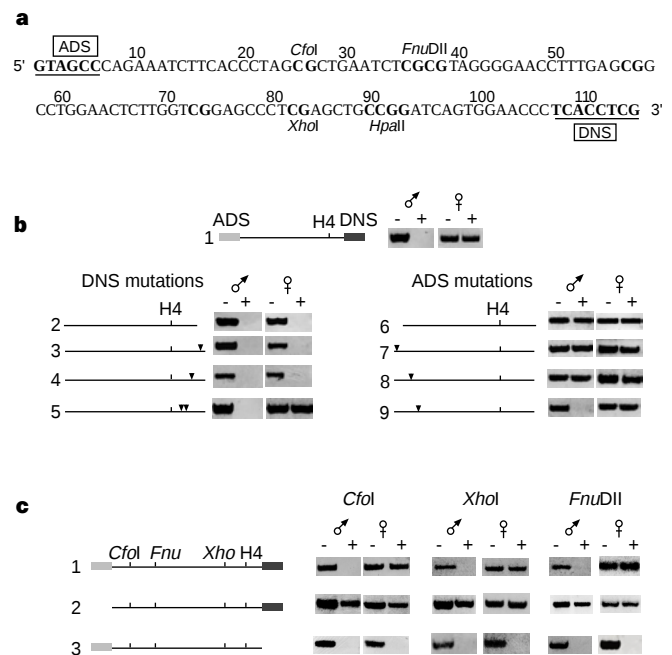


Figure 2 Defining the imprinting box. **a**, The 113-bp sequence comprises the imprinting box. Allele-discrimination signal (ADS) and *de novo* methylation signal (DNS) are in bold and underlined. The *HpaII* site is H4 shown in Fig. 1a. *CfoI*, *FnuDII* and *XhoI* are other CpG-containing sites in the sequence. **b**, All experiments (1–9) were performed by injections of PCR products obtained with primers as specified below. For injection procedure into male and female pronucleus and for methylation analysis, see Methods. Methylation analysis was performed on blastocyst DNA undigested (–) or digested (+) with *HpaII* before the PCR reaction. Experiment 1, injection of the entire imprinting box (primers 7 and 8; see Methods); 2, G at position 113 is deleted (primers 7 and 9); 3, A replaces C at position 112 (primers 7 and 10); 4, T replaces C at position 106 (primers 7 and 11); 5, A replaces C at positions 104, 105 (primers 7 and 12); 6, G in position 1 is deleted (primers 13 and 8); 7, C replaces T at position 2 (primers 14 and 8); 9, T replaces C at position 7 (primers 15 and 8). **c**, Experimental details are as in **b** but the sites analysed are *CfoI*, *XhoI* and *FnuDII*. Experiment 1, the intact imprinting box (primers 7 and 8) was injected; 2, the injected fragment is deleted in ADS (primers 16 and 8); 3, the injected fragment lacks the DNS (primers 7 and 17).

(H4), and we next examined the methylation status of other CpG sites in this region using the same injection experiments. Eight CpG sites are present in the 113-bp sequence, six of which are included in sites recognized by the methyl-sensitive restriction enzymes *CfoI*, *FnuDII*, *XhoI* and *HpaII* (Fig. 2a). To test the effect of ADS and DNS on the methylation of the other CpG sites, we repeated the injection experiments with the intact 113-bp sequence (Fig. 2c, experiment 1), a 107-bp fragment that does not include the 6-bp ADS (Fig. 2c, experiment 2) and a 105-bp fragment that does not include the 8-bp DNS (Fig. 2c, experiment 3). The methylation status of *XhoI*, *CfoI* and *FnuDII* sites in the injected fragments was analysed at the blastocyst stage. We found that methylation of all CpG sites is affected by ADS and DNS in a way similar to that described for site H4.

Next we asked whether ADS and DNS control methylation of the entire 3-kb DMR. We therefore analysed the methylation status of *HpaII* site 3, located 730 bp upstream to site H4. When the 3-kb *PvuII* fragment was injected, site H3 underwent *de novo* methylation only when injected into the female pronucleus. In contrast, no *de novo* methylation of site H3 took place when we removed a 250-bp *Sfi/Mlu* sequence that includes the 113-bp box or when DNS was mutated as described above (Fig. 3a). This result indicates that the monoallelic methylation of site H3 and the establishment of differential methylation of the entire DMR2 is governed by the

elements present in the 113-bp sequence. However, maintenance of this differential methylation throughout the postimplantation stage may require additional control elements not present on this 113-bp fragment. Indeed, using our analytical method, we have observed that this region is methylated in postimplantation embryos, regardless of which pronucleus was injected (data not shown). This result is consistent with previously reported observations derived from *Igf2r* DMR2 transgene experiments⁵ and indicates that the above-mentioned additional control elements are located outside the entire DMR2.

To examine whether these elements can affect methylation of DMRs other than the *Igf2r* DMR, we ligated the 250-bp *Sfi/Mlu* fragment, which includes the 113-bp box, to the 5' region of the mouse imprinted *Snrpn* gene (DMR1)¹². Injection of the 250-bp sequence into the male or female pronucleus of the fertilized egg resulted in the expected maternal methylation of site H4 (Fig. 3b, experiment 1). Injection of the *Snrpn* DMR1 (1.6 kb) did not result in significant *de novo* methylation (Fig. 3b, experiment 2). However, when the 250-bp fragment was ligated to the *Snrpn* DMR1, the *HpaII* site M1 became methylated when injected into the female pronucleus but remained unmethylated when injected into the male pronucleus (Fig. 3b, experiment 3). The 250-bp box did not induce methylation of site M1 of the non-imprinted human *APOA1* gene¹³ (Fig. 3b, experiment 4). These experiments show that the 113-bp

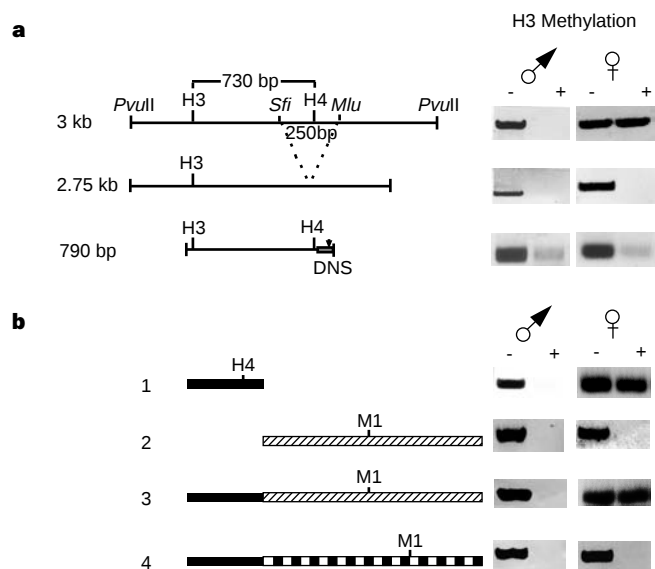


Figure 3 Specificity of signalling of the imprinting box. Experimental details as in Figs 1, 2. **a**, Methylation status of site H3 in blastocysts developed from zygotes injected with the intact DMR2 (3-kb *PvuII* fragment), with a DMR2 deleted in the imprinting box (*Sfi/Mlu* fragment includes the 113-bp imprinting box) and with a 790-bp PCR fragment from DMR2 mutated in DNS (as in Fig. 2b, experiment 3) into the male or female pronucleus. PCR after digestion with *HpaII* (+) or without *HpaII* (-) was with primers 18 and 19 for the intact and the deleted fragments and with primers 18 and 10 for the mutated fragment. **b**, Experiment 1, methylation status of site H4 in blastocysts after injection of the *Sfi/Mlu* 250-bp fragment. Primers for PCR were 1 and 2. Experiment 2, injections were with the *SacI* 1.6-kb fragment, which is part of the *Snrpn* DMR1 (ref. 12) (primers 20 and 21). Experiment 3, the injected fragment was constructed by fusing the 250-bp *Sfi/Mlu* sequence with the *Snrpn* DMR1 (primers 20 and 21). In experiments 2 and 3, the methylation status of M1, a maternally methylated site of *Snrpn*¹², was analysed. Experiment 4, the injected fragment was obtained by ligating the 250-bp *Sfi/Mlu* fragment to the human *APOA1* upstream sequence (see Fig. 1c). In this experiment the upstream *HpaII* site (M1) of the *APOA1* gene¹⁴ was analysed (primers 3 and 4).

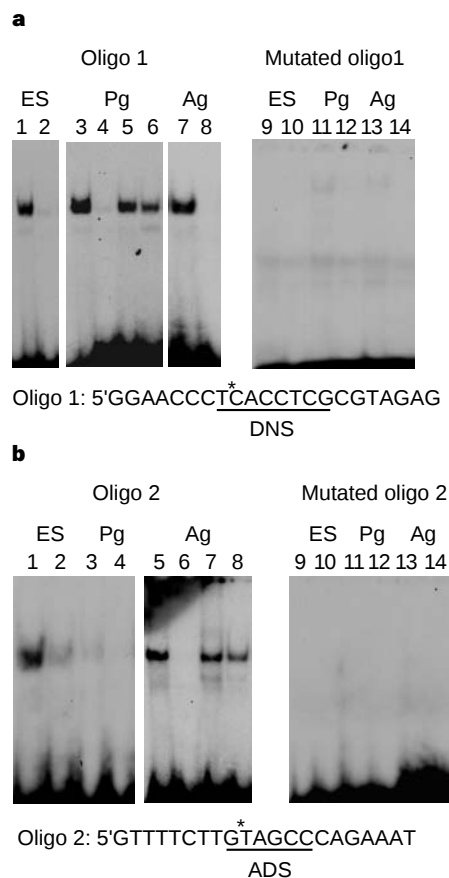


Figure 4 Band-shift assays. Nuclear extracts of ES cells (ES), parthenogenetic ES cells (Pg) and androgenetic ES cells (Ag) were incubated under conditions described (see Methods). **a**, With oligo 1. A 20-bp oligonucleotide that includes the 8-bp DNS sequence, or with oligo 1 mutated at the C residue (*). **b**, With oligo 2. A 20-bp oligonucleotide that includes the 6-bp ADS, or with oligo 2 mutated at the T residue (*). The reactions were done without competitor oligonucleotide (lanes 1, 3, 7, 9, 11, 13 in **a** and 1, 3, 5, 9, 11, 13 in **b**); with 100-fold excess of cold oligonucleotide (lanes 2, 4, 8, 10, 12, 14 in **a** and 2, 4, 6, 10, 12, 14 in **b**); with 100-fold cold mutated oligonucleotide (lane 5 in **a** and lane 7 in **b**); or with 100-fold cold unrelated oligonucleotide (lane 6 in **a** and lane 8 in **b**).

box is capable of inducing monoallelic methylation of a sequence that is recognized as a DMR, but is ineffective when a 'naïve' non-imprinted sequence is presented. Our results also show that whereas *Snrpn* DMR1 has germline-specific methylation *in vivo*¹², it appears to lack its own DNS and ADS, which are provided here by the *Igf2r* fragment.

Having demonstrated that monoallelic methylation of *Igf2r* DMR2 is guided by two *cis* elements within a 113-bp box, we attempted to identify proteins that bind specifically to these elements. In a band-shift assay using nuclear extracts from normal, parthenogenetic and androgenetic embryonic stem (ES) cells a protein binding to a 20-bp oligonucleotide that includes DNS was observed (Fig. 4a, lanes 1, 3, 7). No such band appears when the oligonucleotide was mutated by the same single-base substitutions that abolished *de novo* methylation in the injection experiments (Fig. 4a, lanes 9–14) (methylation of the CpG in DNS did not affect the binding). Binding specificity was corroborated by competition experiments with excess cold oligonucleotide (Fig. 4a, lanes 2, 4, 8) and by failure to compete with mutated (Fig. 4a, lane 5) or unrelated (Fig. 4a, lane 6) oligonucleotides. This specific band shift is evidence for the existence of a nuclear protein (DNP), which may take part in the *de novo* methylation process.

In a band-shift assay exposing nuclear extracts from normal, parthenogenetic or androgenetic ES cells to a 20-bp oligonucleotide that includes ADS, we observed a specific band with normal or androgenetic ES nuclear extract but not with parthenogenetic ES nuclear extract (Fig. 4b). This specific band-shift reflects the existence of a nuclear protein of paternal origin that recognizes this sequence and binds to it specifically. The same single-base substitutions in ADS that interfered with discrimination between the alleles (Fig. 2b) also abolished the binding of this specific protein (Fig. 4b, lanes 9–14). Specificity was again confirmed by competition control experiments.

The protein that binds to ADS (ADP) protects, presumably, the paternal allele from *de novo* methylation (Fig. 5). Because deletion of DMR2 activates the paternal *Igf2r* allele and represses production of paternal antisense RNA, an expression-competition model has been suggested⁵. It is therefore likely that maternal methylation of DMR2 represses production of maternal antisense RNA while paternal antisense RNA can be produced, thereby preventing paternal-specific expression of the gene (Fig. 5).

Imprinting is clearly a multistage mechanism that begins with marking of alleles during gametogenesis, followed by the establishment of DMRs that are maintained by protection from genome-

wide demethylation in the preimplantation embryo, and global *de novo* methylation at the pregastrula stage¹⁴. Here we focused on the establishment of *Igf2r* DMR2, showing that the primary mark which is laid down in the gametes can still be established in the pronucleus upon a short sequence that constitutes the imprinting box. The precise mechanisms and elements involved in the other stages of the imprinting process await elucidation. □

Methods

Preparation of DNA fragments for injection. A 1.8-kb *HindIII/PvuII* upstream fragment of the human apolipoprotein AI (*APOAI*) was obtained from pUC19, which harbours the 5' region of the gene¹³. The 0.8-kb *BamHI/EcoRI* fragment including exon 1 of the human PGK1 was cut out from a pUC19 plasmid that harbours PGK-1 (ref. 15). The 1.6-kb fragment from the mouse *Snrpn* gene was the *SacI* fragment in pUC19 that harbours *Snrpn* DMR1 (ref. 12). The 3-kb *PvuII* fragment was cut out from a pUC9 plasmid that harbours *Igf2r* DMR2 (ref. 7). Shorter sequences around the *HpaII* site in *Igf2r* DMR2 were prepared by PCR with the appropriate primers. The plasmid fragments that were propagated in *Escherichia coli* were methylated *in vivo* at the GATC sites by the bacterial dam methylase, whereas the PCR fragments that were used for the injections were methylated *in vitro* by dam methylase before injection as follows: DNA fragments were incubated overnight at 37 °C with 16 μ M S-adenosylmethionine (SAM) and dam methylase (5–10 units μ g⁻¹ DNA). The DNA was then purified by extraction with phenol/chloroform/isoamyl alcohol (25:24:1 v/v) and ethanol precipitation.

Preparation of zygotes for injection. Four-to-six-week-old (C57Black $\times$ BALB/c) F1 females were superovulated by intraperitoneal injections of 10 units of pregnant mare's serum gonadotropin, followed after 48 h by an injection of 10 units of human chorionic gonadotropin. Superovulated females mated overnight and zygotes were removed from the oviducts 20 h after injection of chorionic gonadotropin. These fertilized eggs were used for the injection experiments.

Injection into zygotes' pronuclei. Approximately 1 μ l of a DNA solution in water (10 ng μ l⁻¹) was injected into the male or female zygote pronucleus. The injected zygotes were incubated at 37 °C, in M16 medium (Sigma) covered with paraffin oil and 5% CO₂/95% air atmosphere to obtain preimplantation embryos at different stages of development. The postimplantation embryos were obtained by transferring injected zygotes into the oviduct of foster mothers. Embryos were dissected from the foster mother's uterus at days 6.5 or 7.5 postcoitum.

Methylation assay of injected fragments. A pool of 10–20 injected embryos was collected; DNA was extracted and subjected to digestion with *DpnI* and *MboI* (pools as small as three to five embryos were also used with similar success). Half of the digested DNA was also subjected to digestion with *HpaII*, in addition to the *DpnI* and *MboI* digestions. *DpnI* cuts DNA at GATC sites when the site is methylated on both strands, whereas *MboI* cuts at GATC when unmethylated on both strands. These two enzymes are refractory to digestion at hemimethylated GATC sites. *DpnI* therefore eliminates unintegrated injected DNA (being episomal, it remains methylated on both strands). *MboI* digests the integrated DNA that underwent at least two rounds of replication, as well as the endogenous counterpart of the injected fragment, as these sequences are unmethylated on both strands. Only hemimethylated molecules that are composed of the original integrated strand (methylated at GATC) and a newly synthesized strand (unmethylated at GATC) should survive digestion by *MboI* and *DpnI* (ref. 11). Because we arranged to have GATC sites adjacent to the studied CpG site, we were able to assay the methylation status of a CpG site on the original integrated DNA. This was done by digestion with a methylation-sensitive restriction enzyme, such as *HpaII*, before amplifications by PCR using primers flanking the assayed CpG site. A PCR product is observed only when the site is methylated and refractory to digestions by the methyl-sensitive restriction enzyme. The PCR products were electrophoresed on 2% agarose gels, stained with ethidium bromide and photographed by the Bio-Imagine System, BIS202 (Dinco & Rhenium Industries, Jerusalem). The PCR primers used for the methylation analysis were: 1, 5'-TCAGAACACTGGTGAGCAGTGGG-3'; 2, 5'-GAGGGTAGGATTCGGTTGCAAGG-3'; 3, 5'-ATGGAC-AATTGGCAACTGCC-3'; 4, 5'-TAAGCAGCCAGCTCTTGCA-3'; 5, 5'-ACCTGGGTCTCGCACATTC-3'; 6, 5'-TGCTGAGCAGCCGCTATTGG-3';

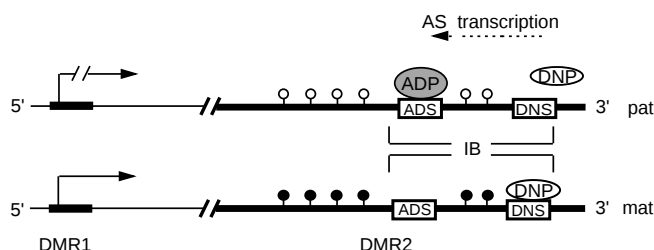


Figure 5 A proposed model for the imprinting of the *Igf2r* gene. The imprinting box (IB), which is located within the intronic differentially methylated region (DMR2), includes two regulatory *cis* elements: the *de novo* methylation signal (DNS) and the allele-discrimination signal (ADS). The protein DNP binds to DNS and guides *de novo* methylation on the maternal allele (mat). The protein ADP binds to ADS on the paternal allele (pat), and prevents binding of DNP to DNS on this allele. As a result, DMR2 becomes methylated on the maternal allele (filled symbols) and remains unmethylated on the paternal allele (open symbols). Paternal transcription of antisense RNA (AS transcription) starts in DMR2 (ref. 17). The full AS transcript may prevent paternal transcription of sense RNA. If methylation of DMR2 on the maternal allele represses AS transcription, it can explain the connection between this methylation and the maternal expression of the *Igf2r* gene.

7, 5'-GTAGCCAGAAATCTTCACC-3'; 8, 5'-CGAGGTGAGGGTTCCTGAT-3'; 9, 5'-GAGGTGAGGGTTCCTGAT-3'; 10, 5'-TACGCTAGGTGAGGGTTCAC-3'; 11, 5'-TACGCGAGGTGGGGTTCCTGAT-3'; 12, 5'-TACGCGAGGTGATTGTTCCA-3'; 13, 5'-TAGCCAGAAATCTTCACC-3'; 14, 5'-GCAGCCAGAAATCTTCACC-3'; 15, 5'-GTAGCCTAGAAATCTTCACC-3'; 16, 5'-CAGAAATCTTCACCCTAGCG-3'; 17, 5'-TGAGGGTTCCTGATCC-3'; 18, 5'-ACCGCAA CTCAGCACAACCAA-3'; 19, 5'-TAGCACAATCTCAATTGTGCTGCG-3'; 20, 5'-CCCTCTCCACATAGTAAAAATCTGT-3'; 21, 5'-CGTCCCAGGCAATGGCTGC-3'.

Band-shift assays. Nuclear extracts were prepared as described before¹⁶ from $3-5 \times 10^7$ ES cells grown in a DMEM:Ham's F-12 medium supplemented with 20 pg ml^{-1} LIF (Gibco). Labelled oligonucleotides were prepared from the double-stranded oligonucleotides listed below by end filling with Klenow DNA polymerase and [α -³²P]dCTP. Labelled oligonucleotides (30–100 pg; 10^4 c.p.m.) were incubated at 30°C for 30 min, with the nuclear extracts (10 µg) in a buffer containing 12 mM HEPES, 60 mM KCl, 0.6 mM Na₂EDTA, 0.6 mM DTT, 5 mM MgCl₂ and 1 µg poly (dI–dC) in a final volume of 20 µl. The reaction mixtures were electrophoresed on a 4% polyacrylamide gel. The normal and mutated (underlined) DNS (in bold letters) double-stranded oligonucleotides were: I, 5'-GGAACCCCT**AC**CTCGCGTAGAG; II, 5'-GGAACCCCT**AC**CT-AGCGTAGAG. The normal and mutated (underlined) ADS (in bold letters) double-stranded oligonucleotides were: I, 5'-GTTTCTT**GT**AGCCAGAAAT; II, 5'-GTTTCTT**GC**AGCCAGAAAT.

Received 4 September; accepted 16 November 1998.

1. Surani, M. A. H. Imprinting and the initiation of gene silencing in the germ line. *Cell* **93**, 309–312 (1998).
2. Razin, A. & Cedar, H. DNA methylation and genomic imprinting. *Cell* **77**, 473–476 (1994).

3. Li, E., Beard, C. & Jaenisch, R. Role for DNA methylation in genomic imprinting. *Nature* **366**, 362–365 (1993).
4. Elson, D. A. & Bartolomei, M. S. A 5' differentially methylated sequence and the 3' flanking region are necessary for H19 transgene imprinting. *Mol. Cell. Biol.* **17**, 309–317 (1997).
5. Wutz, A. *et al.* Imprinted expression of the Igf2r gene depends on an intronic CpG island. *Nature* **389**, 745–749 (1997).
6. Nicholls, R. D., Saitoh, S. & Horsthemke, B. Imprinting in Prader-Willi and Angelman syndromes. *Trends Genet.* **14**, 194–199 (1998).
7. Stoger, R. *et al.* Maternal-specific methylation of the imprinted mouse Igf2r locus identifies the expressed locus as carrying the imprinting signal. *Cell* **73**, 61–71 (1993).
8. Szabo, P. E. & Mann, J. R. Allele-specific expression and total expression levels of imprinted genes during early mouse development: implications for imprinting mechanism. *Genes Dev.* **9**, 3097–3108 (1995).
9. Brandeis, M. *et al.* The ontogeny of allele-specific methylation associated with imprinted genes in the mouse. *EMBO J.* **12**, 3669–3677 (1993).
10. Shemer, R. *et al.* Dynamic methylation adjustment and counting as part of imprinting mechanisms. *Proc. Natl Acad. Sci. USA* **93**, 6371–6376 (1996).
11. Kafri, T., Gao, X. & Razin, A. Mechanistic aspects of genome-wide demethylation in the preimplantation mouse embryo. *Proc. Natl Acad. Sci. USA* **90**, 10558–10562 (1993).
12. Shemer, R., Birger, Y., Riggs, A. D. & Razin, A. Structure of the imprinted mouse Snrpn gene and establishment of its parental-specific methylation pattern. *Proc. Natl Acad. Sci. USA* **94**, 10267–10272 (1997).
13. Shemer, R. *et al.* Methylation changes in the apolipoprotein I gene during embryonic development of the mouse. *Proc. Natl Acad. Sci. USA* **88**, 11300–11304 (1991).
14. Kafri, T. *et al.* Developmental pattern of gene-specific DNA methylation in the mouse embryo and germline. *Genes Dev.* **6**, 705–714 (1992).
15. Keith, D. H., Singer-Sam, J. & Riggs, A. D. Active X chromosome DNA is unmethylated at eight CCGG sites clustered in a guanine-plus-cytosine-rich island at the 5' end of the gene for phosphoglycerate kinase. *Mol. Cell Biol.* **6**, 4122–4125 (1986).
16. Lee, K. A. W., Bindereif, A. & Green, M. R. A small scale procedure for preparation of nuclear extracts that support efficient transcription and pre-mRNA splicing. *Gene Anal. Tech.* **5**, 22–31 (1988).
17. Barlow, D. P. Competition—a common motif for the imprinting mechanism? *EMBO J.* **16**, 6899–6905 (1997).

Acknowledgements. We thank H. Cedar for comments on the manuscript; T. Kafri for help in the initial steps of this project; and B.-Z. Tsuberi for help in the transfer of embryos into foster mothers. The work was supported by NIH, CTR (USA) and Israel Science Foundation.

Correspondence and requests for materials should be addressed to A.R. (e-mail: shemer@md2.huji.ac.il).

YOURS TO HAVE AND TO HOLD BUT NOT TO COPY

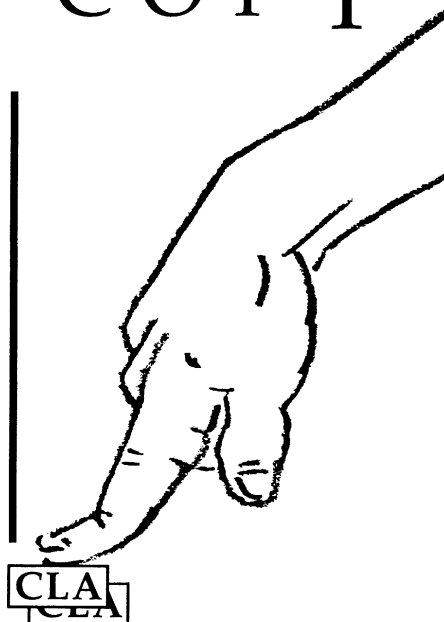
The publication you are reading is protected by copyright law. This means that the publisher could take you and your employer to court and claim heavy legal damages if you make unauthorised infringing photocopies from these pages.

Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

The Copyright Licensing Agency (CLA) is an organisation which issues licences to bring photocopying within the law. It has designed licensing services to cover all kinds of special needs in business, education, and government.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP
Telephone: 0171 436 5931
Fax: 0171 436 3986